FILED UNDER SEAL

COOLEY LLP
BOBBY GHAJAR (198719)
(bghajar@cooley.com)
COLETTE GHAZARIAN (322235)
(cghazarian@cooley.com)
1333 2nd Street, Suite 400
Santa Monica, CA 90401
Telephone: (310) 883-6400

KATHLEEN HARTNETT (314267)
(khartnett@cooley.com)
MARK WEINSTEIN (193043)
(mweinstein@cooley.com)
JUDD LAUTER (290945)
(jlauter@cooley.com)
ELIZABETH L. STAMESHKIN (260865)
(lstameshkin@cooley.com)
3175 Hanover Street
Palo Alto, CA 94304
Telephone: (650) 843-5000

CLEARY GOTTLIEB STEEN & HAMILTON LLP
ANGELA L. DUNNING (212047)
(adunning@cgsh.com)
1841 Page Mill Road, Suite 250
Palo Alto, CA 94304
Telephone: (650) 815-4131

PAUL, WEISS, RIFKIN, WHARTON & GARRISON LLP
KANNON K. SHANMUGAM (*pro hac vice*)
(kshanmugam@paulweiss.com)
2001 K Street, NW
Washington, DC 20006
Telephone: (202) 223-7300

[Full Listing on Signature Page]
Counsel for Defendant Meta Platforms, Inc.

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. 3:23-cv-03417-VC-TSH

DEFENDANT META PLATFORMS, INC.'S: (1) NOTICE OF MOTION AND MOTION FOR PARTIAL SUMMARY JUDGMENT; AND (2) OPPOSITION TO PLAINTIFFS' MOTION FOR PARTIAL SUMMARY JUDGMENT

Date: May 1, 2025
Time: 10:00 a.m.
Dept: Courtroom 4 – 17th Floor
Judge: Vince Chhabria
Trial Date: None
Date Action Filed: July 7, 2023

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page(s)

Cases

Statutes

Other Authorities

TABLE OF EXHIBITS

EX. NO.	DESCRIPTION
EXHIBITS TO DECLARATION OF BOBBY GHAJAR ("BG")	
1	Excerpts of Plaintiffs' deposition testimony, each testifying as to whether he or she is aware of any Llama outputs that replicate text from his or her works
2	Excerpts of Plaintiffs' deposition, each testifying as to whether he or she is aware of Llama outputs that are substantially similar to or that might substitute for Plaintiffs books
3	Excerpts of Plaintiffs' Responses to Meta's Requests for Admissions, namely Responses to Request No. 24 (No. 26 as to Pl. Farnsworth) in which each admitted, subject to objections, that he or she is not personally aware of any text generated by any of Meta's Llama models that infringes his or her at-issue works
4	Excerpts from Plaintiff Klam's deposition, testifying that Llama would not generate text from his books in response to his prompting
5	Excerpts of Plaintiffs' Responses to Meta's Requests for Admissions, namely Responses to Request Nos. 22–23 (Nos. 24–25 at to Pl. Farnsworth), which concern each Plaintiffs' awareness, if any, of any instance in which a person read text generated by any of Meta's Llama models as a substitute for reading his or her at–issue work(s) or any documentary evidence of the same
6	Excerpts of certain Plaintiffs' deposition testimony, each testifying as to their use, if any, of generative AI tools
7	Excerpts of certain Plaintiffs' deposition testimony, each testifying as to the purposes of generative AI as compared to Plaintiffs' books
8	Excerpts of certain Plaintiffs' deposition testimony, each testifying as to whether he or she is aware of any lost sales or licensing opportunities due to the infringement alleged in this case
9	Excerpts of Plaintiffs' Responses to Meta's Requests for Admissions, namely Responses to Request Nos. 12, 15, and 18, concerning Plaintiffs' awareness of any lost sales or licensing opportunities due to the infringement alleged in this case or of any documentary evidence of the same
10	Excerpts of certain Plaintiffs' deposition testimony, each testifying as to whether he or she has ever received an offer to license, ever offered to license, and ever actually licensed his or her works to train AI
11	Excerpts of Plaintiffs' Responses to Meta's Requests for Admissions, namely Responses to Request Nos. 8–10 and 31, which concern Plaintiffs' licensing of their at–issue works for use as training data for artificial intelligence

EX. NO.	DESCRIPTION
12	Excerpts of Plaintiffs' Responses to Meta's Interrogatories, namely Responses to Interrogatory No. 24, concerning the identity of third parties with whom Plaintiffs have discussed the possibility of licensing any of their at–issue books as training data for large language models
13	Excerpts of certain Plaintiffs' deposition testimony, each testifying as to whether Llama affected his or her intent to continue writing books
14	Excerpts of certain Plaintiffs' deposition testimony, each testifying as to whether he or she wrote his or her books with an expectation or intent to monetize his or her works through licenses for AI training
15	Excerpts of certain Plaintiffs' deposition testimony, each testifying as to whether he or she would license their books for AI training
16	Excerpts of certain Plaintiffs' deposition testimony, each testifying as to his or her opinions on generative AI
17	Excerpts of certain Plaintiffs' deposition testimony, each testifying as to whether his or her publisher(s) have the right to license his or her books for AI training
18	Excerpts of Plaintiffs' Responses to Meta's Requests for Admissions, namely Responses to Request Nos. 69 and 70 (as to Pl. Coates); 67 and 68 (as to Pl. Díaz); 71 and 72 (as to Pl. Golden); 70 and 71 (as to Pl. Greer); 69 and 70 (as to Pl. Hwang); 75 and 76 (as to Pl. Kadrey); 67 and 68 (as to Pl. Klam); 75 and 76 (as to Pl. Lippman); 65 and 66 (as to Pl .Silverman); 65 and 66 (as to Pl. Snyder); 69 and 70 (as to Pl. TerKeurst); 83 and 84 (as to Pl. Woodson); and 74 and 75 (as to Pl. Farnsworth), concerning the rights, if any, of the publishers of Plaintiffs' at–issue works to license those works as training data for large language models
19	Excerpts of Plaintiffs' Responses to Meta's Requests for Admissions, namely Responses to Request Nos. 73 and 74 (as to Pl. Coates); 71 and 72 (as to Pl. Díaz); 78 and 79 (as to Pl. Farnsworth), 75 and 76 (as to Pl. Golden), 74 and 75 (as to Pl. Greer), 73 and 74 (as to Hwang), 79 and 80 (as to Pl. Kadrey), 71 and 72 (as to Pl. Klam), 79 and 80 (as to Pl. Lippman), 69 and 70 (as to Pl. Silverman), 69 and 70 (as to Pl. Snyder), 73 and 74 (as to Pl. TerKeurst), and 87 and 88 (as to Pl. Woodson), concerning the published status of Plaintiffs' at–issue works
20	Excerpts from the transcript of the deposition of Plaintiffs' expert Dr. Emily Bender, taken February 25, 2025
21	Excerpts from the transcript of the deposition of Plaintiffs' expert Dr. Jonathan Krein, taken March 6, 2025
22	Excerpts of Exhibit 1 (Opening Expert Report of Dr. Jonathan Krien) to the deposition of Plaintiffs' expert Dr. Jonathan Krein, taken March 6, 2025, together

EX. NO.	DESCRIPTION
	with relevant excerpts from the deposition transcript in which the exhibit was introduced and authenticated
23	Excerpts from the transcript of the deposition of Plaintiffs' rebuttal expert Dr. Cristina Lopes, taken February 13, 2025
24	Excerpts from the transcript of the deposition of Plaintiffs' rebuttal expert Dr. Cristina Lopes, taken March 7, 2025
25	Excerpts from the transcript of the deposition of Plaintiffs' expert Dr. Daniel Spulber, taken March 3, 2025
26	Excerpts from the transcript of the Rule 30(b)(1) deposition of Amrish Acharya, taken November 21, 2024
27	Exhibit 643 to the Rule 30(b)(1) deposition of Amrish Acharya, taken November 21, 2024, together with relevant excerpts from the deposition transcript in which the exhibit was introduced and authenticated
28	Intentionally omitted
29	Excerpts from the transcript of the Rule 30(b)(1) deposition of Meta engineer Nikolay Bashlykov, taken December 5, 2024
30	Excerpts from the transcript of the Rule 30(b)(1) deposition of research scientist Sean Bell, taken December 11, 2024
31	Excerpts from the transcript of the Rule 30(b)(1) deposition of Alexander Boesenberg, taken November 18, 2024
32	Excerpts from the transcript of the Rule 30(b)(1) deposition of Sy Choudhury, taken December 5, 2024
33	Excerpts from the transcript of the Rule 30(b)(6) deposition of Meta, by and through its corporate designee, Sy Choudhury, taken December 5, 2024
34	Excerpts from the transcript of the Rule 30(b)(1) deposition of Sy Choudhury, taken January 14, 2025
35	Excerpts from the transcript of the Rule 30(b)(1) deposition of Michael Clark, taken November 13, 2024
36	Excerpts from the transcript of the Rule 30(b)(1) deposition of Michael Clark, taken November 14, 2024
37	Excerpts from the transcript of the Rule 30(b)(6) deposition of Meta, by and through its corporate designee, Michael Clark, taken December 19, 2024

EX. NO.	DESCRIPTION
38	Excerpts from the transcript of the Rule 30(b)(6) deposition of Meta, by and through its corporate designee, Michael Clark, taken March 3, 2025, on the topic of alleged torrenting
39	Intentionally omitted
40	Excerpts from the transcript of the Rule 30(b)(1) deposition of Sergey Edunov, taken November 6, 2024
41	Excerpts from the transcript of the Rule 30(b)(1) deposition of David Esiobu, taken December 13, 2024
42	Excerpts from the transcript of the Rule 30(b)(1) deposition of Melanie Kambadur, taken September 17, 2024
43	Excerpts from the transcript of the Rule 30(b)(1) deposition of Dr. Yann LeCun, taken November 21, 2024
44	Excerpts from the transcript of the Rule 30(b)(1) deposition of Chaya Nayak, taken October 31, 2024
45	Excerpts from the transcript of the Rule 30(b)(1) deposition of Eugene Nho, taken December 6, 2024
46	Excerpts from the transcript of the Rule 30(b)(1) deposition of Joelle Pineau, taken November 6, 2024
47	Excerpts from the transcript of the Rule 30(b)(1) deposition of Thomas Scialom, taken December 5, 2024
48	Excerpts from the transcript of the Rule 30(b)(1) deposition of Hugo Touvron, taken December 3, 2024
49	Excerpts from the transcript of the deposition of Meta's expert Professor Lyle Ungar, PhD, taken February 26, 2025
50	Excerpts of Meta's Response to Plaintiffs' Interrogatories, namely its Responses to Interrogatory No. 1, concerning data used to train the Llama models
51	Excerpts of Meta's Response to Plaintiffs' Interrogatories, namely its Responses to Interrogatory No. 7, concerning, among other things, Meta's efforts to mitigate the risk of the Llama models reproducing training data verbatim
52	Copy of a document produced by Meta with a first Bates number Meta_Kadrey_00000185 (Leo Gao et al., *The Pile: An 800GB Dataset of Diverse Text for Language Modeling*, EleutherAI (Dec. 31, 2020))

EX. NO.	DESCRIPTION
53	Plaintiffs' Supplemental Initial Disclosures, served December 13, 2024
EXHIBIT TO DECLARATION OF PROF. MICHAEL SINKINSON, PH.D. ("SINKINSON")	
A	Curriculum Vitae of Prof. Michael Sinkinson, Ph.D.
EXHIBIT TO DECLARATION OF PROF. LYLE UNGAR, PH.D. ("UNGAR")	
A	Curriculum Vitae of Prof. Lyle Ungar, Ph.D.
EXHIBIT TO DECLARATION OF BARBARA FREDERICKSON-CROSS ("FREDERICKSON")	
A	Curriculum Vitae of Barbara Fredericksen-Cross
EXHIBITS TO DECLARATION OF CHAYA NAYAK ("NAYAK")	
A	Copy of document produced by Meta, Bates numbered Meta_Kadrey_00000078–Meta_Kadrey_00000104 (Hugo Touvron et al., *LLaMa: Open and Efficient Foundation Language Models*, METAAI (Feb. 2023)), cited in Plaintiffs' Motion for Partial Summary Judgment (Dkt. 472) at 7 n.15
B	*Introducing LLaMA: A foundational, 65-billion-parameter large language model*, Meta blog (Feb. 24, 2023), https://ai.meta.com/blog/large-language-model-llama-meta-ai/
C	*Celebrating 1 Billion Downloads of Llama*, Meta (Mar. 18, 2025), https://about.fb.com/news/2025/03/celebrating-1-billion-downloads-llama/
D	Copy of document produced by Meta, Bates numbered Meta_Kadrey_00149022–Meta_Kadrey_00149027 (LLaMA 1 License Agreement)
E	Copy of document produced by Meta, Bates numbered Meta_Kadrey_00235561–Meta_Kadrey_00235563 (LLaMA 2 Community License Agreement (July 18, 2023))
F	Copy of document produced by Meta, Bates numbered Meta_Kadrey_00000157–Meta_Kadrey_00000159 (Meta Acceptable Use Policy)
G	Copy of document produced by Meta, Bates numbered Meta_Kadrey_00168426–Meta_Kadrey_00168502 (Hugo Touvron et al., *Llama 2: Open Foundation and Fine-Tuned Chat Models*, GenAI, Meta (July 2023))
H	Copy of document produced by Meta, Bates numbered Meta_Kadrey_00093669–Meta_Kadrey_0093760 (Llama Team, *The Llama 3 Herd of Models* (July 23, 2024))
I	*Llama 3.2: Revolutionizing edge AI and vision with open, customizable models*

EX. NO.	DESCRIPTION
	(Sept. 25, 2024), https://ai.meta.com/blog/llama-3-2-connect-2024-vision-edge-mobile-devices/
J	Copy of document produced by Meta, Bates numbered Meta_Kadrey_00093662–Meta_Kadrey_00093665 (Meta Llama 3 Community License Agreement (Apr. 3, 2024))
K	Copy of document produced by Meta, Bates numbered Meta_Kadrey_00093242–Meta_Kadrey_0093253 (*Meet Zoom AI Companion, your new AI assistant! Unlock the benefits with a paid Zoom account* (updated July 22, 2024))
L-1	*With 10x growth since 2023, Llama is the leading engine of AI innovation*, Meta blog (Aug. 29, 2024), https://ai.meta.com/blog/llama-usage-doubled-may-through-july-2024/
L-2	*How Organizations Are Using Llama to Solve Industry Challenges*, Meta blog (Jan. 13, 2025), https://about.fb.com/news/2025/01/organizations-using-llama-solve-industry-challenges/
M	*Meditron: An LLM suite especially suited for low-resource medical settings leveraging Meta Llama*, Meta blog (Apr. 25, 2024), https://ai.meta.com/blog/llama-2-3-meditron-yale-medicine-epfl-open-source-llm/
N	Copy of document produced by Meta, Bates numbered Meta_Kadrey_00093619–Meta_Kadrey_00093621 (Envision webpage)
O	*Discover the possibilities of building on Llama*, Meta (Mar. 19, 2025), https://www.llama.com/community-stories/
P	*Llama Impact Grants*, Meta (Mar. 19, 2025), https://www.llama.com/llama-impact-grants/
DECLARATION OF NIKOLAY BASHLYKOV ("BASHLYKOV")	
	No exhibits
DECLARATION OF DAVID ESIOBU ("ESIOBU")	
	No exhibits

PLEASE TAKE NOTICE that on May 1, 2025, at 10:00 a.m., in Courtroom 4 of the above Court, Defendant Meta Platforms, Inc. will and hereby does move, under Fed. R. Civ. P. 56, for an order granting summary judgment to Meta on grounds that: (1) Meta's copying of Plaintiffs' works to develop and train large language models is fair use and, thus, does not infringe Plaintiffs' rights under 17 U.S.C. § 106(1); and (2) Meta did not violate 17 U.S.C. § 1202. Plaintiffs' Motion for Partial Summary Judgment (Dkt. 472) ("Plaintiffs' Motion") should be denied in full.

MEMORANDUM OF POINTS AND AUTHORITIES

I. INTRODUCTION

This case presents a question of existential importance to the future of generative artificial intelligence ("AI") development in the United States: whether Meta's copying of text from the Internet, including copyrighted books, to develop and train large language models ("LLMs") capable of generating an astonishing array of new and non-infringing content constitutes fair use under U.S. copyright law. The answer, as a matter of law, should be "yes."

Meta's LLMs, known as Llama, are an extraordinary technology capable of providing human-like responses to user "prompts." To achieve this functionality, Llama was "trained" on trillions of "tokens" consisting of individual words and letter combinations found in myriad data, including source code, websites, books, scientific articles, and other sources. This process extracts and converts grammatical, syntactical, and other patterns and information from the training corpus into billions of numerical "weights" that define relationships between all tokens in the datasets. After being trained on vast and varied data, LLMs can use the building blocks of language in remarkable ways, including to "generate creative text, solve mathematical theorems, predict protein structure, answer reading comprehension questions, and more. They are one of the clearest cases of the substantial potential benefits AI can offer at scale to billions of people." Nayak, Ex. B.

Although other developers have trained and offer proprietary LLMs, Meta released the code and weights for its first Llama model, Llama 1, in early 2023 on a non-commercial basis to academic researchers, governmental organizations, and industry labs with the goal of facilitating development of better LLMs and more innovative use cases. Meta released subsequent open-source versions of Llama for both research and commercial purposes subject to permissive licenses.

Today, millions of individuals, researchers, and businesses use Llama, or their own platforms built on Llama, *for free* to innovate and deliver new technologies, products, and services to the public.

Plaintiffs are the authors of 49 works included in the datasets used to train Llama. They do not claim that Llama has ever reproduced their works or could serve the same purpose as their novels, plays, and memoirs: one cannot read those works using Llama. They admit they have no evidence that Llama has cost them sales. Nevertheless, Plaintiffs assert that Meta should have sought and paid for permission to copy their works to train Llama, and that its failure to do so infringed their copyrights. *See* Pls. Mot. at 3, 21.[1] This does not square with U.S. copyright law.

The Copyright Act has "never accorded the copyright owner complete control over all possible uses of his work." *Sony Corp. of Am. v. Univ. City Studios, Inc.*, 464 U.S. 417, 432 (1984) ("*Sony Betamax*").[2] Where a "use is otherwise fair, then no permission need be sought or granted." *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 585 n.18 (1994). This "fair use" doctrine, codified in the Copyright Act of 1976, developed to "avoid the rigid application of the copyright statute when, on occasion, it would stifle the very creativity which that law is designed to foster." *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 18 (2021) ("*Oracle*"). Fair use "reflects a balance of competing claims upon the public interest: Creative work is to be encouraged and rewarded, but private motivation must ultimately serve the cause of promoting broad public availability of literature, music, and other arts." *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 526 (2023) ("*Goldsmith*"). Thus, copyright protection is not an end unto itself, but a means "to further copyright's very purpose, to promote the Progress of Science and useful Arts." *Authors Guild v. Google, Inc.*, 804 F.3d 202, 212 (2d Cir. 2015) ("*Google Books*") (cleaned up).

Courts balance four non-exclusive factors to assess whether a use is fair: (1) nature of the

[1] Only 11 of the 13 Plaintiffs have moved for summary judgment. Pls' Mot. at 3 n.4. They seek a determination that Meta's "*copying*" of their works infringed their *reproduction* rights under 17 U.S.C. § 106(1), and that the "initial *acquisition*" of copies of those works "cannot be fair use." *Id.* at 2; *see id.* at 20–21. Plaintiffs do not seek summary judgment on their *distribution* claim under § 106(3), which is premised on allegations that Meta also uploaded copies of their works to others while downloading them. Meta's motion is directed to why its copying to develop and train Llama was fair use and not infringing. Meta has separately sought leave to submit a rebuttal expert report directed to Plaintiffs' newly added *distribution* claim (Dkt. 486), and, upon completion of the newly ordered discovery, may request leave to seek summary adjudication of that claim, if warranted.

[2] Unless otherwise noted, all emphasis is added and internal citations and quotation marks are omitted.

use; (2) nature of the copyrighted work; (3) substantiality of the portion used; and (4) effect of the use upon the potential market for or value of the original. 17 U.S.C. § 107. These factors must be applied flexibly to embrace "significant changes in technology." *Oracle*, 593 U.S. at 19. Here, these factors—especially the most important first and fourth factors, weigh heavily in favor of fair use.

Factor One: It is difficult to imagine a technology more transformative than Llama. A use is transformative where it "adds something new, with a further purpose or a different character," rather than "merely 'supersed[ing] the objects' of the original creation." *Goldsmith*, 598 U.S. at 528 (quoting *Campbell*, 510 U.S. at 579). Llama is nothing like a book; it is not meant to be read. Instead, Llama is a *tool* that predicts and generates original content in response to user queries based on statistical modeling derived from its training. Llama can serve as a personal tutor on nearly any subject, assist with creative ideation, and help users to generate business reports, translate conversations, analyze data, write code, and compose poems or letters to friends. It can answer questions about authors or help readers find books they might enjoy. What it does *not* do is replicate Plaintiffs' books or substitute for reading them. Llama has a distinct and quintessentially transformative purpose, weighing decisively in favor of fair use.

Factor Two: The second factor also favors Meta. Plaintiffs' works are all published, and the statistical information Meta extracted from them to create its Llama models is not protectable.

Factor Three: This factor also favors Meta, because the copying was reasonable relative to its purpose. Plaintiffs acknowledge that Llama could not exist absent training on extensive data, and it has never outputted more than a few, short passages from some of Plaintiffs' books—far less than what was held to be fair use as a matter of law in *Google Books*.

Factor Four: Plaintiffs cannot show that Llama has adversely affected them, let alone in a manner cognizable under copyright law. They admit they have no evidence of lost sales. They admit they did not write their books with any inkling they could be used to train an LLM. Indeed, the only opportunity they claim to have lost due to Meta's use of their works to train Llama is to license and charge Meta for that exact, unintended use—a "circular[]" argument rejected in *Oracle*. 593 U.S. at 38. Plaintiffs have never licensed their books for AI training; no such market exists even today, years after Llama's release; and copyright law does not allow authors to monopolize

markets for transformative uses of their works in any event. This lack of harm weighs heavily in favor of fair use here, particularly when balanced against Llama's immense public benefits.

Generative AI has ushered in tremendous capabilities across an expanding array of use cases and industries. Llama, alone, has been downloaded by more than a *billion* individuals and businesses, enabling scientific discovery and new means of creating non-infringing expression: the very definition of promoting the progress of science and the arts. The public interest, and the purposes of copyright, would be badly disserved by precluding Meta from making transformative use of copyrighted text to build cutting-edge AI technology that does not substitute for Plaintiffs' books. No court has ever declined to find fair use in such circumstances.

In sum, Meta's copying of datasets containing Plaintiffs' books was for a transformative fair use and "not an infringement of copyright." 17 U.S.C. § 107. Further, the undisputed facts show that Meta did not remove copyright management information from Llama's training data with intent to conceal infringement in violation of 17 U.S.C. § 1202. The Court should deny Plaintiffs' Motion in full and enter judgment for Meta on Plaintiffs' §§ 106(1) and 1202 claims.

II. SUMMARY OF UNDISPUTED FACTS

A. Plaintiffs

Named plaintiffs Richard Kadrey, Sarah Silverman, Christopher Golden, Ta-Nehisi Coates, Junot Díaz, Andrew Sean Greer, David Henry Hwang, Matthew Klam, Laura Lippman, Rachel Louise Snyder, Jacqueline Woodson, Lysa TerKeurst, and Christopher Farnsworth (together, "Plaintiffs") are 13 book authors who claim to own registered copyrights in the 49 books identified in Exhibit A to the Third Amended Complaint ("TAC") (Dkt. 407). These works include novels (*e.g.*, Klam's *Who is Rich?*), plays (*e.g.*, Hwang's *M. Butterfly*), and a memoir (*e.g.*, Silverman's *The Bedwetter*), all general audience "trade books." Sinkinson ¶¶ 13–14. These books were included in the extensive datasets that Meta copied to train its Llama family of LLMs. Pls' Mot. at 15, 19–21. Plaintiffs Coates and Golden have not moved for summary judgment on any claim. *Id.* at 3 n.4; *see* Pritt App'x A, Ex. 1 (excluding their 7 works).

B. Large Language Models ("LLMs") and Generative AI

Artificial intelligence dates to the mid-20th century, when early researchers set out to create

machines that simulate human intelligence. Ungar ¶ 10. The first AI systems were developed to perform actions like playing chess largely based on pre-programmed logical rules. *Id*. In parallel, and picking up in the 1980s, researchers also explored the concept of "neural networks," an alternative approach to AI inspired by the human brain in which computer systems learn rules from patterns in data on which they are "trained," instead of relying on pre-programmed instructions. *Id*. ¶ 11. Neural networks consist of interconnected nodes (or "neurons") arranged in layers that progressively extract increasingly complex mathematical patterns from training data, enabling the network to output a prediction or decision based on the patterns derived. *Id*. ¶ 14.

LLMs are an advanced "deep learning" neural network designed to understand and generate text by analyzing contextual relationships between "tokens" in training data. *Id*. ¶¶ 15–29. Tokens are the basic units of text processed by LLMs, consisting of words, word segments, and punctuation (e.g., "The Golden Gate Bridge is red." has 7 tokens including the period). *Id*. ¶ 19. LLMs encode information derived from tokens in "parameters," numerical "weights" that determine the strength of connections between nodes. *Id*. ¶¶ 12, 31–38. During the "pretraining" process, vast quantities of tokens are entered into the model to adjust the weights based on complex training steps repeated billions or trillions of times. *Id*. ¶ 32. Training datasets are necessarily "diverse" and "immense because LLMs begin with random weights and must learn everything about language from scratch—word meanings, syntax, structure, and world knowledge—from statistical patterns in text alone." *Id*. ¶ 44; BG Ex. 23 (Lopes 2/13 Dep.) at 53:18–54:9. Once so trained, LLMs can generalize across many types of contexts by predicting which token is most likely to follow the last in a given sequence, allowing them to generate original text responses to a virtually infinite array of user prompts. Ungar ¶¶ 34–36; BG Ex. 23 (Lopes 2/13 Dep.) at 39:20–40:17. Many models also undergo "post-training" (or "finetuning") in order to align them to specific tasks or objectives. Ungar ¶ 30.

The November 2022 release by OpenAI of its proprietary ChatGPT model brought public attention to the immense power and promise of LLMs. Sinkinson ¶ 10. But Meta and other developers had long been working on their own models (*id*.; BG Ex. 46 (Pineau Dep.) at 44:20–45:2), and in February 2023, Meta released its first flagship LLM, Llama. Nayak ¶ 6. Unlike OpenAI, Meta made its model available to researchers to download and experiment on at no cost. *Id*. ¶¶ 15–

16. Others, including Google and Anthropic, also soon released advanced proprietary LLMs of their own, establishing the United States as the global leader in AI innovation. Sinkinson ¶¶ 10, 78.

C. Meta and Its Llama Models

Meta: Meta has long been a leader in developing technology to connect people, including via Facebook, Instagram, and WhatsApp. For over a decade, it has also been at the forefront of AI research and development. In 2013, Meta launched the Fundamental AI Research ("FAIR") lab under the leadership of Yann LeCun, a Turing Award-winning pioneer in deep learning and neural networks. BG Ex. 43 (LeCun Dep.) at 118:14–21, 267:12–268:1. In early 2023, Meta reorganized its AI research teams, creating a new division, GenAI. Nayak ¶ 4. This restructuring separated FAIR's foundational AI research from GenAI's product-driven innovation, enabling Meta to advance both cutting-edge research and practical applications in parallel. *Id*.

Llama 1: Llama development began in 2022 as a research project focused on creating an LLM to assist with proving mathematical theorems. BG Ex. 46 (Pineau Dep.) at 51:1–53:1. Using an early Google LLM as a benchmark (*id*. at 119:1–11; Ex. 48 (Touvron Dep.) at 124:8–125:2), Meta pretrained Llama 1 on a mix of publicly available datasets comprising computer code, scientific papers, books, and miscellaneous content from the Internet. Nayak ¶¶ 6–12 & Ex. A at 2. Combined, these datasets included 1.4 *trillion* tokens. *Id*. By contrast, a typical book is 50,000 to 150,000 tokens, or no more than 0.0000107% of the Llama 1 training corpus. *See* Ungar ¶ 60.

One of the datasets Meta used, commonly referred to as "Books3," contained the text of more than 190,000 books, including Plaintiffs' at-issue works. Nayak ¶ 10; Pls.' Mot. at 7. Meta obtained content from Books3 via direct download from a third-party website. BG Ex. 22 ¶¶ 71–74. Books3 was assembled and published by a third party AI researcher in 2020 to promote AI development. Ungar ¶ 48. Books3 was also released as a part of "The Pile," a larger text dataset compiled by non-profit research group EleutherAI. Dkt. 407 ¶¶ 40–41, 45; BG Ex. 52. Over 100 organizations, including technology companies, universities, and non-profits, have published research using the Pile and Books3, which collectively have been cited over 10,000 times. Ungar ¶ 48. Plaintiffs' technical expert agreed at her deposition that "open-source dataset[s]" like the Pile are "important to the advancement of LLM research." BG Ex. 23 (Lopes 2/13 Dep.) at 58:12–15.

Books3 was not the only source of books used to train Llama 1. Meta also used Project Gutenberg, which contains more than 70,000 books in the public domain. Dkt. 407 ¶ 39. Books3 and Project Gutenberg, *combined*, made up 4.5% of the Llama 1 training data. Nayak ¶ 10. The vast majority of data used to train Llama 1 came from Common Crawl, while other data sources came from sources like Wikipedia and Github (a repository of open source code). *Id.* ¶¶ 7–12.

Meta initially released Llama 1 in February 2023 on a noncommercial basis to academics, governmental organizations, and research labs around the world, with the goal of democratizing and advancing AI research. *Id.* ¶¶ 6, 13 & Ex. B. The model was released in four sizes—7 billion, 13 billion, 33 billion, and 65 billion parameters. Ungar ¶ 40.

Llama 2: In July 2023, shortly after this action was filed, Meta released a second Llama model—Llama 2—in three sizes (7 billion, 13 billion, and 70 billion parameters), along with a research paper detailing its architecture and training process. *Id.*; Nayak ¶ 17 & Ex. G. Llama 2 was pretrained on the same datasets as Llama 1, including Books3, but also underwent extensive finetuning to improve the safety, quality, and consistency of model outputs. *Id.* ¶ 14; BG Ex. 51. It was released on an open-source basis for both research and commercial use. Nayak ¶ 16.

Llama 2 represented a major contribution to AI research and development. Designing and training LLMs is time consuming and expensive. Ungar ¶ 73. By releasing Llama 2 under a permissive, open-source license, Meta enabled start-ups, non-profits, and researchers to learn from, experiment with, and iterate on state of the art AI to which they otherwise would not have access. *Id.* ¶¶ 70–80. And they did. The Llama 2 paper has been cited over 11,000 times. *Id.* ¶ 74.

Llama 3: Between April and December 2024, Meta released multiple versions of Llama 3, delivering significant improvements in performance and efficiency, while maintaining its commitment to open access and broad industry adoption. Nayak ¶¶ 18–20, 22. Ranging from 1 billion to 405 billion parameters, the Llama 3 models allow developers to balance size relative to performance, with smaller models being less powerful but also less computationally intensive and capable of running on mobile devices. *Id.* & Ex. I. Llama 3's release also marked a significant expansion in Meta's use of Llama for its own services, including the launch of Meta AI, a general purpose AI chatbot, as a standalone web application in April 2024. *Id.* ¶ 23; Sinkinson ¶ 20.

Llama 3 was pretrained on a significantly expanded, diverse corpus of more than *15 trillion* tokens. Ungar ¶ 43 ("Llama 3 dataset is so large that if printed onto standard letter-sized paper, it would produce a stack of paper more than 930 miles high, which is approximately the distance between Los Angeles and Seattle."); Nayak ¶ 22; BG Ex. 50 (listing datasets). Consistent with the increase in overall training data volume, Meta downloaded additional books data from a website known as "Library Genesis" or "Libgen" in Spring 2023. Bashlykov ¶ 4. During this period, the only portion of Libgen downloaded via a torrent protocol was "SciMag," which contains academic publications (but not Plaintiffs' books). Bashlykov ¶¶ 4–5; Pls' Mot. at 13. Meta used a direct download method to copy books data from two other portions of Libgen, "Fiction," and "Scitech," which contained some of Plaintiffs' books. Bashlykov ¶ 4; BG Ex. 22 ¶¶ 107–16.

Llama 4: Llama 4 is a larger, more advanced LLM planned for release this year. BG Ex. 44 (Nayak Dep.) at 60:5–61:15. To obtain the additional tokens necessary to train it, Meta downloaded, via torrent, books data from a third party repository called "Anna's Archive" ("AA"). BG Ex. 38 (Clark Dep.) at 52:2–14; Ex. 45 (Nho Dep.) at 118:1-16; Ex. 30 (Bell Dep.) at 97:7–21, 133:14–17 (Llama 4 to be trained on 30 to 60 trillion tokens). AA aggregates links to Libgen and Z/Library (or Z/Lib), and books from the Internet Archive. *Id.* at 64:3–12. As with the torrent from SciMag described above, Meta used a torrenting script for AA that prevents seeding (i.e., distribution after file download). BG Ex. 38 (Clark Dep.) at 52:2–14, 93:1-6, 121:17–22; Bashlykov ¶ 5.

For all Llama models, the objective was to acquire large volumes of text for training, not any particular work. Bashlykov Decl. ¶ 6; BG Ex. 23 (Lopes 2/13 Dep.) at 20:19-22 ("Q: And in order to pretrain a large language model, does that require a large amount of data? A: Yes"); Ex. 25 (Spulber Dep.) at 277:13-16 (Plaintiffs' expert, Prof. Spulber: "the training and development of AI models, such as Meta's Llama models, requires extraordinarily large amounts of data"); Ex. 40 (Edunov Dep.) at 51:22–52:5; 54:20–55:2.[3] Because any given work is a tiny fraction of total training data (for Llama 3, conservatively, only 0.000001% assuming a book of average length), no individual text materially contributes to performance of the model. Ungar ¶¶ 60–64. It is

[3] Plaintiffs emphasize the number of copies of their works Meta downloaded. Pls' Mot. at 2, 15, 25. Multiple copies of many works—for some, dozens—were included in large datasets like Z/Lib. Fredricksen ¶ 22 n.17. Meta deduplicated these datasets before training, as described below.

through analyzing the aggregation of trillions of words across millions of texts that useful patterns, structures, and linguistic relationships emerge. *Id*. ¶¶ 42–47.

D. Llama Does Not Replicate Plaintiffs' Works

Plaintiffs admit that they are not aware of any Llama outputs that replicate their text. BG Ex. 1. Several Plaintiffs testified that they use LLMs for, among other things, inspiration for their writing, recipes, medical diagnoses, and research, but agreed Llama does not output their works. *Id*. ¶ 8 & Ex. 5. Nor does Llama generate substantially similar outputs that might substitute for Plaintiffs' books. No Plaintiff could identify any example of this. *Id*. ¶ 2 & Ex. 4 (excerpts). Several conceded that Llama serves a different purpose from their books. *Id.* ¶ 9 & Ex. 7 (same).

Plaintiffs assert that Llama can generate outputs that "cop[y] significant portions" of their books, Pls' Mot. at 17 n.29, but this is unsupported. Referred to as "memorization," LLMs can sometimes reproduce short segments of training data duplicated in the training corpus. Ungar ¶ 49. "Memorization" is a misnomer, however, as unlike data retrieved from fixed memory (e.g., a hard drive), LLM outputs are based on probabilistic predictions. *Id*. When an LLM trains on the same text sequence repeatedly, it may "overfit" on that text, i.e., overestimate the probability of that sequence. *Id*. ¶¶ 49–50. Thus, LLMs sometimes "memorize" short segments of oft-quoted texts such as the U.S. Constitution, which appears many times in datasets such as Common Crawl. *Id.*

Beginning with Llama 2, Meta undertook a number of mitigations to reduce memorization risk. BG Ex. 36 (Clark 11/13 Dep.) at 52:14–53:11; Ex. 37 (Clark 12/19 Dep.) at 25:10–36:10; Ex. 51. This included deduplicating training data and, during finetuning, teaching models to refuse to respond to prompts for potentially copyrighted content. BG Ex. 47 (Scialom Dep.) at 160:16–161:20, Ex. 48 (Touvron Dep.) at 256:13–258:7); Esiobu ¶ 6. Plaintiff Klam demonstrated the effectiveness of this mitigation; he asked Llama to output text from his book, only to receive a response stating it would not do so. BG Ex. 4. He also confirmed he could not identify any language from his books generated by Meta AI. BG Exs. 1, 4.

Meta conducted experiments to ensure that memorization rates were low prior to release; they were very low. *Id*. Ex. 37 (Clark Dep.) at 30:14–31:13; Esiobu ¶¶ 3–5. This was confirmed by Meta's expert, Dr. Lyle Ungar, who conducted experiments showing that, on average, Llama

could continue passages from Plaintiffs' books by only a few words. Ungar ¶¶ 51–59. Plaintiffs' expert, Dr. Lopes, was able to coax the models into reproducing several sentences of some books by inputting book passages as prompts and asking the model to continue them. BG Ex. 23 (Lopes 2/13 Dep.) at 141:22–143:5. There is no evidence the public uses Llama in this way and, in any case, Dr. Lopes was able to generate less than a paragraph of text. *Id.* 179:22–181:8; Ex. 24 (Lopes 2/14 Dep.) at 237:16–19 ("Q: You are not offering any opinion that Llama is able to reproduce, you know, any significant percentage of these books, correct? A: Correct.").

Relatedly, as Dr. Ungar explains, the training process transforms training data into something completely new and unrecognizable. Ungar ¶ 38; BG Ex. 20 (Bender Dep.) at 187:2–11, 193:23–194:6. As text passes through a neural network's layers, "billions of weights are iteratively adjusted" to produce a mathematical representation of the relationship between every word in the training corpus, but nothing of the training data itself remains. Ungar ¶ 18.

Given its transformative nature and functionality, Llama's release has not cost Plaintiffs any book sales or licensing opportunities in any normal markets (e.g., as screenplays, audiobooks). Sinkinson ¶¶ 12, 58 (discussing the original market for Plaintiffs' books and traditional subsidiary rights). Plaintiffs admit they are unaware of any evidence of such harm. BG Exs. 8, 9. To validate this, Meta's expert, Dr. Sinkinson, conducted a regression analysis using publicly available book sales data to test whether Llama had any effect on Plaintiffs' book sales, and found it had none. Sinkinson ¶¶ 18–35. Plaintiffs' damages expert, Dr. Spulber, speculates in his report that Llama outputs could, at some point, compete with Plaintiffs' books for readers' attention. BG Ex. 25 (Spulber Dep.) at 263:18–266:10. But he acknowledged this had not occurred (*id*. 245:7–248:1), that he's currently unaware of any instance in which Meta's LLM has substituted for Plaintiffs' works (*id.* at 274:2–248:1), in which a book has been written using Llama, or in which someone has read a Llama output in lieu of a book (*id.* at 263:18–266:10).

E. Llama Greatly Benefits the Public

Llama has become a cornerstone of open-source AI innovation, enabling researchers and businesses to experiment with and build upon Meta's efforts and achievements at no cost. Nayak ¶¶ 24–32; Ungar ¶¶ 70–80; Sinkinson ¶¶ 72–75. The Llama models have been downloaded more

than a billion times since their launch (Nayak ¶ 15), and the papers introducing Llama 1 and 2 have been cited in more than 7,000 research publications. Ungar ¶ 74. Meta has also made Llama available to all government agencies and several government partners. Sinkinson ¶ 76. Generative AI has ushered in tremendous capabilities across many industries, contributing billions to our economy and prompting warnings from successive administrations against disadvantaging U.S. companies and the public by allowing foreign states to pull ahead in AI development. *Id.* ¶¶ 77–80.

Oracle, ScaleAI, and Lockheed Martin are all using Llama to develop national security programs and to supplement existing data analysis and code generation functions. *Id*. Yale School of Medicine has partnered with a Swiss institute to develop Meditron, an open-source LLM built on Llama to improve access to evidence-based information for clinical decision making. Nayak ¶ 28 & Ex. M. Envision, an assistive learning technology company, integrates Llama's language processing and computer vision into a mobile app and prototype glasses that can translate visual information into speech. *Id*. ¶ 29 & Ex. N. Non-profit Jacaranda Health is using Llama to provide personalized health support in Swahili to Kenyan mothers. *Id.* ¶ 31 & Ex. P. And popular meeting platform Zoom has leveraged Llama to develop tools that enable users to transcribe and summarize video meetings. *Id*. ¶ 26 & Ex. K. *See also id*. ¶¶ 27, 30, 31 & Exs. L, O, P (additional use cases).

III. ARGUMENT

A. Legal Standard for Summary Judgment

Summary judgment is proper when "there is no genuine dispute as to any material fact and the movant is entitled to judgment as a matter of law." Fed. R. Civ. P. 56(a). Fair use is a mixed question of law and fact. *Oracle*, 593 U.S. at 24. "Where no material, historical facts are at issue and the parties dispute only the ultimate conclusions to be drawn from those facts, [the court] may draw those conclusions without usurping the function of the jury." *Seltzer v. Green Day, Inc.*, 725 F.3d 1170, 1175 (9th Cir. 2013). Courts routinely grant summary judgment for copyright defendants on fair use grounds. *See, e.g., Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 822 (9th Cir. 2003); *Google Books*, 804 F.3d at 230; and numerous other cases cited below.

B. Overview of Core Copyright and Fair Use Principles

The root of U.S. copyright law is the English Statute of Anne, 8 Anne, C.19 (1710), which

sought "to encourage creativity and ensure that the public would have free access to information by putting an end to 'the continued use of copyright as a device of censorship.'" *Suntrust Bank v. Houghton Mifflin Co.*, 268 F.3d 1257, 1260 (11th Cir. 2001). The Constitution's Copyright Clause authorizes Congress to pass laws "to promote the Progress of Science and the useful Arts" by "securing for limited Times to Authors ... the exclusive Right to their respective writings." U.S. Const. art. 1, § 8, cl. 8. The Copyright Act of 1790, 1 Stat. 124, like the Statute of Anne, was called "An Act for the Encouragement of Learning." *See Google Books*, 804 F.3d at 212 n.11. The "ultimate goal of copyright is to expand public knowledge and understanding[.]" *Id.* at 212.

"The Copyright Act encourages creativity by granting to the creator of an original work 'a bundle of rights.'" *Goldsmith*, 598 U.S. at 509; *see* 17 U.S.C. § 106 (conferring six "exclusive rights," including the right to "reproduce the copyrighted work in copies," § 106(1), and to "distribute copies … to the public," § 106(3)).[4] These rights are afforded not "based upon any natural right that the author has in his writings, but upon the ground that the welfare of the public will be served and progress of science and useful arts … promoted" thereby. *Sony Betamax*, 464 U.S. at 429 n.10. "[R]eward to the owner [is] a secondary consideration" in service to the "primary object" of benefitting the public. *United States v. Paramount Pictures, Inc.*, 334 U.S. 131, 158 (1948).

Copyright's objectives also "parallel those of the First Amendment." Pierre N. Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev. 1105, 1110 (1990) ("Leval").[5] Copyright is meant "to be the engine of free expression." *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 558 (1985); *Sony Betamax*, 464 U.S. at 431 n.12 (copyright has "always been closely connected with freedom of expression"). To this end, protection is constrained by three important limitations.

First, copyright is given for a limited term, ensuring that works pass into the public domain when copyrights expire. *Suntrust Bank*, 268 F.3d at 1262.

---

[4] Copyright protection "extends only as far as Congress designates by statute." *Thaler v. Perlmutter*, 2025 WL 839178, at *1 (D.C. Cir. Mar. 18, 2025) (citing *Wheaton v. Peters*, 33 U.S. 591, 661 (1834)). Further, copyright ownership "is distinct from ownership of any material object in which the work is embodied." 17 U.S.C. § 202. One must thus distinguish "between ownership of the work, which an author does not possess [apart from a specific copy], and ownership of the copyright [in the work], which an author enjoys for a limited time." *Suntrust Bank*, 268 F.3d at 1263.

[5] The Supreme Court relied on Judge Leval's article in *Campbell*, *Oracle*, and *Goldsmith*, as do Plaintiffs (*see* Pls' Mot. at 23), and it is considered a seminal authority on modern fair use analysis.

Second, copyright protects original expression, not facts or ideas. A "fundamental axiom of copyright law is that no author may copyright his ideas or the facts he narrates." *Feist Publ'ns., Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 344–45 (1991); 17 U.S.C. § 102. This principle, known as the "idea/expression dichotomy," dictates that "every idea, theory, and fact in a copyrighted work becomes instantly available for public exploitation at the moment of publication." *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003); *Oracle*, 593 U.S. at 18 ("copyrights protect 'expression' but not the 'ideas' that lie behind it"). And it is critical to the advancement of science and art. *See Nash v. CBS, Inc.*, 899 F.2d 1537, 1540 (7th Cir. 1990) ("Intellectual (and artistic) progress is possible only if each author [is free to] build[] on the work of others."); *Campbell*, 510 U.S. at 575 ("Every book in literature, science and art … must necessarily borrow[] and use much which … [came] before.").

Third, copyright law allows fair use of even protected expression. "From the infancy of copyright protection, some opportunity for fair use … has been thought necessary to fulfill copyright's very purpose[s]," *Campbell*, 510 U.S. at 575, which include both "promoting broad public availability of literature, music, and the [] arts," *Goldsmith*, 598 U.S. at 526, and "preservation of a meaningful public or democratic dialogue," *Suntrust Bank*, 268 F.3d at 1263. "Fair use is not just excused by the law, it is wholly authorized by the law." *Lenz v. Univ. Music Corp.*, 815 F.3d 1145, 1151–52 (9th Cir. 2016). Anyone "may reproduce a copyrighted work for a 'fair use'; the copyright owner does not possess the exclusive right to such a use," and anyone "who makes a fair use of the work is not an infringer … with respect to such use." *Sony Betamax*, 464 U.S. at 433. By excluding transformative uses from the ambit of copyright, the fair use doctrine serves as a "context-based check that can help to keep a copyright monopoly within its lawful bounds" by ensuring authors are afforded no "more economic power than is necessary to achieve the incentive to create" and cannot "us[e] copyright to stifle innovation" by others. *Oracle*, 593 U.S. at 21–22.

C. Meta's Copying of Plaintiffs' Works to Train Llama Was Fair Use, Warranting Judgment for Meta on Plaintiffs' § 106(1) Claim

Meta does not dispute that it made copies of datasets containing Plaintiffs' works to train Llama. But such copying is a quintessential fair use—*not* infringement. This brief first addresses why the fair use doctrine authorizes Meta's copying as a matter of law, warranting summary

judgment in its favor on Plaintiffs' § 106(1) claim. It then addresses Plaintiffs' misguided fair use arguments, which would have the Court ignore Meta's actual "use" in assessing whether it was fair.[6]

> The fair use doctrine, codified in Section 107 of the 1976 Copyright Act, provides:
>
> [T]he fair use of a copyrighted work, ... for purposes such as criticism, comment, news reporting, teaching ... scholarship, or research, is *not an infringement of copyright*. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—
>
> (1) the purpose and character of the use, including whether such use is of a commercial nature …;
>
> (2) the nature of the copyrighted work;
>
> (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
>
> (4) the effect of the use upon the potential market for or value of the copyrighted work.

17 U.S.C. § 107. No factor is dispositive, *Goldsmith*, 598 U.S. at 527, and the "list of factors is not exhaustive," *Oracle*, 593 U.S. at 19. Likewise, the enumerated fair uses (*e.g.*, criticism, research) are "illustrative, [] not limitative," *Campbell*, 510 U.S. at 577—all involve "use of an original work to serve a manifestly different purpose." *Goldsmith*, 598 U.S. at 528. Section 107 "set[s] forth general principles, the application of which requires judicial balancing, depending upon relevant circumstances, including 'significant changes in technology.'" *Oracle*, 593 U.S. at 19; *see also* H.R. Rep. 94-1476, pp. 65–66 (1976) ("courts must be free to adapt the [fair use] doctrine … on a case-by-case basis," "especially during a period of rapid technological change"). A use is fair if it "serves the copyright objective of stimulating productive thought and public instruction without excessively diminishing the incentives for creativity." *Leval*, at 1110.

Here, Meta's copying of books to train Llama furthers the purposes of copyright by enabling the creation of a transformative new technology that serves a manifestly different purpose from Plaintiffs' books without any cognizable harm to Plaintiffs. The Court should find that this copying constitutes non-infringing fair use as a matter of law.

1. The purpose and character of the use strongly favor fair use (Factor One)

It would be inconsistent with the purposes of copyright to allow the limited monopoly

[6] This motion is directed to the claims of all 13 named Plaintiffs before the Court. *Corbin v. Time Warner Ent.-Advance/Newhouse P'ship*, 821 F.3d 1069, 1085 (9th Cir. 2016). Meta has additional, individualized defenses with respect to the standing of certain Plaintiffs (including but not limited to Coates and Golden, who have not sought summary judgment against Meta) and certain of Plaintiffs' individual works, all of which are expressly reserved.

conferred by Plaintiffs' copyrights to interfere with the development of a new technology as innovative and quintessentially transformative as Llama. As the Court predicted,[7] Plaintiffs neither acknowledge, nor attempt to refute, this unavoidable conclusion in their Motion.

a. Llama and its capabilities are highly transformative

Under the first factor, the "purpose and character of the use," courts examine whether the new use is "transformative"—i.e., whether it "merely 'supersede[s] the objects' of the original creation … or instead adds something new, with a further purpose or different character." *Goldsmith*, 598 U.S. at 527–298 (quoting *Campbell*, 510 U.S. at 579). "[A] use that has a distinct purpose is justified because it furthers the goal of copyright … without diminishing the incentive to create." *Id.* at 510–11; *Google Books*, 804 F.3d at 214 ("The more the appropriator [uses] copied material for new, transformative purposes, the more it serves copyright's goal of enriching public knowledge and the less likely [] the appropriation will serve as a substitute for the original[.]").

Llama is radically transformative. At the most basic level, an LLM is nothing like a trade book. Unlike books, which consist of expressive text meant to be read, Llama consists of software and numerical weights containing no expressive text at all. It does not merely "add" something new; it is an entirely new technology designed to generate novel, context-driven responses to a vast array of user queries. Ungar ¶¶ 33–36. If prompted, some Llama outputs may reflect information about a book or help users find it. Sinkinson ¶¶ 36–54 (explaining how LLMs can enhance book discoverability); BG Ex. 4. But Plaintiffs do not claim that anyone can read or reproduce their books using Llama. *Id.* ¶ 3, Ex. 1. As Plaintiffs' expert stated, the Llama models are "different things" from Plaintiffs' books. *Id.* ¶ 22, Ex. 20; *see also* Dkt. 56 at 1 ¶ 1 (holding: "There is no way to understand the LLaMA models themselves as a recasting … of the plaintiffs' books.").

Time and again, courts have held uses far less transformative than Meta's to be fair at summary judgment. We start with *Google Books*, as it is the most factually analogous case and was repeatedly cited with approval on factor one in *Goldsmith*, 598 U.S. at 531, 545, 548. In *Google Books*, the Second Circuit had "no difficulty concluding that Google's making of a digital

[7] *See* Sept. 20, 2024 Hrg. Tr. at 8:18-24 (Court: "Well, transformative use, I mean there's not going to be a dispute about that, is there?").

copy of [millions of] Plaintiffs' books for the purpose of enabling a search for identification of books containing a term of interest to the searcher involve[d] a highly transformative purpose[.]" 804 F.3d at 216. This was so even though up to 16% of each book could, in fact, be read using Google's tool. *Id.* at 224. The purpose of Google's copying was "to make available significant information about those books," not to reproduce them. *Id.* at 217. So, too, here, Meta extracted information about the word usage in Plaintiffs' books (and other texts) to develop a new technology that serves a distinct purpose and does not permit users to read even one page from Plaintiffs' books.

The Ninth Circuit's *Kelly* decision is also instructive. There, the defendant offered a search engine that enabled users to search for images. 336 F.3d at 815. To create this technology, Arriba used a crawler to copy images from websites without authorization, including Kelly's copyrighted photographs, which it stored and displayed as "thumbnail" images in response to user queries. *Id.* The Ninth Circuit affirmed summary judgment for defendant, finding that this was a transformative fair use and reasoning that "Arriba's use of the images serves a different function than Kelly's use—improving access to information on the internet versus artistic expression." *Id.* at 819.[8]

In *Author's Guild v. HathiTrust*, the Second Circuit also affirmed summary judgment, finding that copying millions of books to create a searchable database was "quintessentially transformative" because "the result of a word search is different in purpose, character, expression, meaning, and message from the page (and the book) from which it is drawn." 755 F.3d 87, 97 (2d Cir. 2014). And in *A.V. ex rel. Vanderhye v. iParadigms, LLC*, the Fourth Circuit affirmed summary judgment, finding that copying entire papers to create a plagiarism detector was "transformative" as the use "was completely unrelated to expressive content[.]" 562 F.3d 630, 634, 640 (4th Cir. 2009).

The Supreme Court's post-trial decision in *Oracle* is also on point. After noting that "the ultimate 'fair use' question primarily involves legal work," 593 U.S. at 24, the Court held that the "purpose and character" of Google's copying of thousands of lines of Sun Java code "was

[8] *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007), reached the same conclusion. Citing *Kelly*, the court held Google's display of thumbnail images in search results transformative and fair because "a search engine provides social benefit by incorporating an original work into a new work, namely, an electronic reference tool," which "provides an entirely new use." *Id.* at 1165.

transformative" and supported a fair use finding, *id.* at 31–32. This use, it reasoned, "seeks to create new products … to expand the use and usefulness of Android-based smartphones. [Google's] new product offers programmers a highly creative and innovative tool for a smartphone environment … that could be readily used by programmers," and, as such, "its use was consistent with that creative 'progress' that is the basic constitutional objective of copyright itself." *Id.* at 30. Meta's Llama is even more innovative, transformative, and useful, and fulfills that same fundamental objective.[9]

And in *Sony Computer Entertainment, Inc. v. Connectix Corp.*, the Ninth Circuit reversed a preliminary injunction, finding that copies of Sony code made to reverse-engineer the PlayStation platform "were protected fair use." 203 F.3d 596, 599 (9th Cir. 2000) ("*Connectix*"). Although competitive, the purpose of the copying was to "create[] a new platform," the Virtual Game Station, which did not "itself contain object code that infringes Sony's copyright." *Id.* at 606. Thus, the court was "at a loss to see how Connectix's drafting of entirely new object code for its VGS program could not be transformative, [even] despite the similarities in function and screen output." *Id.* at 606–07. Here, Meta made copies of Plaintiffs' books to train Llama on statistical information about their language and syntax *without* including any protected expression in its code or weights. That information is then used to enable Llama to perform functions and create outputs completely unrelated to, and different from, reading Plaintiffs' books. Llama is, thus, precisely the kind of "highly creative and innovative tool" the Supreme Court found to be "consistent with that creative 'progress' that is the basic constitutional objective of copyright itself." *Oracle*, 593 U.S. at 30.

b. Llama's commercial availability does not undermine the transformative nature of Meta's use

"Transformative" works are "at the heart of the fair use doctrine's guarantee of breathing

[9] In discussing factor three, *Oracle* noted that the 11,500 lines of code copied were only 0.4% of the entire API at issue, weighing in favor of fair use. 593 U.S. at 3–4. However, courts also find, in assessing factor *one*, that a use is more likely transformative where the original work is an "inconsequential portion" of the accused work, because it serves to show how different the new work is. *See Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 611 (2d Cir. 2006) (finding fair use at summary judgment); *Solid Oak Sketches, LLC v. 2K Games, Inc.*, 449 F. Supp. 3d 333, 348 (S.D.N.Y. 2020) (finding fair use at summary judgment; tattoos were "inconsequential portion of NBA 2K" game, "only appear[ed] on three out of 400 available players," constituted less than 0.000431% of total game data, and "cannot be seen clearly during gameplay"). Here, any single book comprised a tiny fraction of Llama's training data—approximately 0.0000107% for Llama 1, and even less for Llama 3. Ungar ¶ 60. This only further underscores Llama's transformativeness.

space within the confines of copyright," and the "more transformative the new work, the less will be the significance of other factors, like commercialism[.]" *Campbell*, 510 U.S. at 578. "Many of the most universally accepted forms of fair use, such as news reporting and commentary, … as well as parody, are all normally done commercially for profit." *Google Books*, 804 F.3d at 219. Thus, the Supreme Court has often held that for-profit uses are fair. *See Oracle*, 593 U.S. at 32 (Android platform, though commercial, was transformative and fair); *Campbell*, 510 U.S. at 594 (commercial song parody was fair); *see also Google Books*, 804 F.3d at 218–19 (commercial purpose of the Google Books database did not outweigh its "highly transformative" purpose); *Blanch v. Koons*, 467 F.3d 244, 248, 253 (2d Cir. 2006) (defendant paid $2 million for art work held to be fair use).

Here, Llama 1 was released solely for *non*-commercial research purposes, and later versions are available *for free* in most cases. Nayak ¶¶ 15–16. Nevertheless, Meta acknowledges that it is a commercial enterprise, that Llama is used for both commercial and non-commercial purposes, and that Meta hopes one day to recoup its significant investment in this important new technology. BG Ex. 26 (Acharya Dep.) at 338:4–19, 381:14–383:3, 430:2–433:10. Under the above authorities, the first fair use factor still weighs strongly for Meta given Llama's highly transformative purpose.

c. Meta's copying of Plaintiffs' works to train Llama is transformative and fair no matter how it acquired the works

Plaintiffs' Motion is almost entirely devoted to castigating Meta for copying, via both direct download and torrent, large, third party datasets that Meta did not create but "knew contained pirated works." Pls' Mot. at 17.[10] Plaintiffs argue that "copying entire works from pirated databases to avoid compensating the rights holder cannot be fair use," and that, "for fair use to apply, the work that was copied must have been lawfully acquired." *Id.* at 19, 22. Such arguments are not only unsupported—they directly contravene Supreme Court precedent.

Every case of alleged copyright infringement involves an unauthorized use of the plaintiff's work. Often, a defendant often makes fair use knowing the copyright holder would not consent. *See Campbell*, 510 U.S. at 572 (finding fair use even though the plaintiff had "refused permission");

[10] Plaintiffs cite numerous documents out of context and without a sponsoring witness. Although these documents have no bearing on Meta's fair use defense, Meta reserves all objections to them.

Oracle, 593 U.S. at 38 (Google's copying fair even though its licensing discussions with Oracle failed). As the Court explained in *Oracle*, "skepticism about whether bad faith has *any* role in a fair use analysis" is "justifiable, as '*[c]opyright is not a privilege reserved for the well-behaved*.'" 593 U.S. at 33 (citing *Campbell*, 510 U.S. at 585 n.18; *Leval*, at 1126). Thus, *Oracle* questioned whether good faith is even "a helpful inquiry" and noted that "the strength of the other factors pointing toward fair use" rendered evidence of bad faith "not determinative[.]" 593 U.S. at 32–33.

Courts recognize, after *Oracle*, that the concept of good or bad faith has "little influence, one way or the other," on the fair use analysis. *See Teradyne, Inc. v. Astronics Test Sys., Inc.*, 2023 WL 9284863, at *17 n.17 (C.D. Cal. Dec. 6, 2023) (finding fair use at summary judgment), *aff'd*, 2025 WL 341828 (9th Cir. Jan. 30, 2025); *Santos v. Kimmel*, 745 F. Supp. 3d 153, 165 (S.D.N.Y. Aug. 19, 2024) (dismissing on fair use grounds and noting: "bad faith is not dispositive of the fair use question, or even of the first factor," which "still favors defendants in light of the transformative nature of the secondary use"), *appeal pending*, No. 24-2196 (2d Cir.); *Thomson-Reuters Enter. Centre GmbH v. Ross Intel. Inc.*, 2025 WL 458520, at *8 (D. Del. Feb. 11, 2025) ("[e]ven if relevant, bad faith would not move the needle") (citing *Oracle*).

This is consistent with copyright's objectives. "Copyright protection is not withheld from authors who lie, cheat, or steal to obtain their information." Leval, at 1126. This has "no bearing" on copyright protection, because "[c]opyright is not a reward for goodness but a protection for the profits of activity that is useful to the public education." *Id.* Likewise, fair use "focus[es] not on the morality of the secondary user, but on whether her creation claiming the benefits of the doctrine is of the type that should receive those benefits," "with a primary focus on whether the secondary use is productive and transformative and whether it causes excessive injury to the market for the original." *Id.* As the Court stated in *Goldsmith*, "fair use is an *objective* inquiry into *what a user does with an original work*, not an inquiry into [his or her] *subjective* intent." 598 U.S. at 512.

Under *Oracle* and *Goldsmith*, it does not matter whether Meta downloaded datasets containing "pirated" books from a third-party who lacked authorization to distribute them, or borrowed 49 used books from the library and scanned them by hand to achieve the same result. Its use was undeniably transformative, and any attempt to overcome that inevitable conclusion with

rhetoric about Meta's subjective mindset—even if credited—is unavailing as a matter of law.

All of Plaintiffs' cited cases predate *Oracle* and *Goldsmith*, and none alters the analysis here. To begin, *Harper & Row*, is fully consistent with a fair use finding here. 471 U.S. at 562. In that case, President Ford had licensed rights to publish his memoir and prepublication excerpts to Harper & Row and Time Magazine, respectively. The Nation, however, "purloined" a copy of an *unpublished* manuscript and published an article that included core verbatim excerpts—13% of its text—with the "intended purpose of supplanting the copyright holder's commercially valuable right of first publication" by "scooping" Time. *Id.* The Court noted that "[f]air use presupposes good faith," and held that usurping an opportunity to commercialize an unpublished memoir was not fair. *Id.* However, it did not hold that bad faith (or use of an unauthorized copy) is dispositive of the first factor, or of fair use more broadly. If it had, there would have been no need for the Court to analyze in depth all four factors as it did. Instead, the Court focused on the fact that the use was not transformative (the public could read large portions verbatim), the manuscript was unpublished (weighing heavily against fair use), and the Nation article intentionally destroyed the value of Time's prepublication rights—all facts readily distinguishable from those here. *Id.* at 562–63.

In *Los Angeles News Service v. KCAL-TV Channel 9*, the defendant obtained an unauthorized copy of video footage from another station after it was denied a license from the copyright owner, copied the "most valuable" 30 seconds, superimposed its own logo on the footage, and used it in a 4 minute 40 second news broadcast, i.e., "the same purpose" for which LANS had already licensed it to others. 108 F.3d 1119, 1121–23 (9th Cir. 1997). Yet, the Ninth Circuit did not find that using an unauthorized copy precluded fair use. To the contrary, it held this was "not dispositive" and remanded to the trial court to balance the fair use factors. *Id.* at 1122–23.[11]

The only case Plaintiffs cite for the proposition that a party invoking fair use must use an "authorized copy" of a work is *Atari Games Corp. v. Nintendo of America, Inc*., 975 F.2d 832 (Fed. Cir. 1992), which does not help them either. *Atari* acquired Nintendo software source code both

[11] The same was true in *Perfect 10*. Although the Ninth Circuit nodded to the "good faith and fair dealing underpinnings of the fair use doctrine" cited in *Harper & Row*, it held that Google's fair use defense was likely to succeed and rejected Perfect 10's argument that providing access to infringing websites constitutes bad faith that is "inherently not fair use." 508 F.3d at 1164 n.8

by reverse engineering Nintendo chips and by lying to the U.S. Copyright Office to obtain it. *Id.* at 836. The Federal Circuit found that Atari's reverse engineering to build its own, competing program, though unauthorized, was fair use, while use of the "purloined" source code was not. *Id.* at 843–44. Without analysis, the court cited *Harper & Row* for the proposition that "[k]nowing exploitation of a purloined manuscript [is] not compatible with [the] 'good faith' … underpinnings of fair use" and extrapolated a rule—nowhere announced in *Harper & Row*—that "[t]o invoke the fair use exception, an individual must possess an authorized copy of a literary work." *Id.* at 843.

Plaintiffs' singular reliance on *Atari* betrays the fundamental weakness of their argument. *Atari* is not good law. It is an outdated, out-of-circuit case that misstates the holding of *Harper & Row* and has been superseded by *Campbell*, *Oracle* and *Goldsmith*.[12] It is also distinguishable on its facts.[13] Plaintiffs cannot dodge a fair use finding by asserting that the copies Meta made to train Llama were "unauthorized." Even assuming that to be true, factor one (like all other factors), weighs strongly in favor of fair use given Llama's indisputable transformativeness.

2. The "nature of the copyrighted work" favors fair use (Factor Two)

The second factor encompasses consideration of both the type of work copyrighted and its publication status. First, works that are "creative in nature," *Perfect 10*, 508 F.3d at 1167, are "closer to the core of intended copyright protection." *Kelly*, 336 F.3d at 820. However, information *about* creative works, like the facts they contain, are not protectable at all. 17 U.S.C. § 102; *Feist*, 499 U.S. at 344–45. Thus, in *Connectix*, the Ninth Circuit held that copying software code to access its unprotected functional elements was fair use. 203 F.3d at 599. Key to the analysis was that "if Connectix was to gain access to the functional elements of the" software system at issue, "it had to be through a form of reverse engineering that required copying the [system at issue] onto a computer." *Id.* at 603. Permissible copying also extends to "those expressive elements of the work

[12] Even before *Oracle*, other circuits refused to follow *Atari*. *See, e.g., NXIVM Corp. v. Ross Inst.*, 364 F.3d 471, 478–79, 482 (2d Cir. 2004) (assuming defendant had knowingly used an "unauthorized" copy of an unpublished manuscript but still finding that its "transformative secondary uses" were fair and refusing to follow *Atari*, reasoning: "nothing in *Harper Row* indicates that the defendants' bad faith is itself conclusive of the fair use question, or even of the first factor." Moreover, "after *Campbell*, it is clear that a finding of bad faith, or a finding on any one of the four factors, cannot be considered dispositive.") (cleaned up).

[13] To train Llama, Meta copied third party datasets containing Plaintiffs' works from publicly available websites; there is no evidence that it lied to obtain those copies, as in *Atari*.

that must necessarily be used as incident to expression of the underlying ideas, functional concepts, or facts." *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1524 (9th Cir. 1992) (cleaned up).

Second, the fact a copyrighted work was "already published at the time of the copying weighs in favor of fair use." *In re DMCA § 512(h) Subpoena to Twitter, Inc.*, 608 F. Supp. 3d 868, 880 (N.D. Cal. 2022) (citing *Kelly*); *Kelly*, 336 F.3d at 820 ("Published works are more likely to qualify as fair use because the first appearance of the artist's expression has already occurred.").

Here, all of the books at issue were previously published and made available to the public. BG Ex. 19. Further, although Plaintiffs' works are creative, the aspects of the works that Meta needed to extract and use to train Llama are unprotected statistical data regarding word order, frequencies, grammar, and syntax, *i.e.*, unprotectable information *about* Plaintiffs' use of language (relative to all other training data), rather than their protected expression. Ungar ¶¶ 19–29 (describing the process of an LLM deriving statistical relationships between words from training). The necessity to access unprotected aspects of works as to which Plaintiffs had already "exploited [the] commercially valuable right of first publication," *Perfect 10*, 508 F.3d at 1167, tips the second factor in favor of fair use. *See Google Books*, 804 F.3d at 220, 225 (finding second factor weighed for fair use where published books were copied to extract information about the words they contain, not for their expression, and separately noting that copyright "does not include an exclusive right to furnish the kind of information about the works that Google's programs provide to the public").

3. Meta's copying of entire books was reasonable (Factor Three)

The third factor asks whether, as here, the amount copied "[is] reasonable in relation to the purpose of the copying." *Campbell*, 510 U.S. at 586; *Oracle*, 593 U.S. at 33 (copying is fair where "central to a copier's valid purpose"). "[T]he extent of permissible copying varies with the purpose and character of the use." *Campbell,* 593 U.S. at 586–87. *Google Books* further refined the inquiry, noting that what matters is "not so much 'the amount and substantiality of the portion used' *in making a copy*, but rather the amount and substantiality of *what is thereby made accessible* to a public for which it may serve as a competing substitute." 804 F.3d at 222 (emphasis in original).

Courts routinely find that copying an entire work is fair where reasonable or necessary to achieve the purpose of the fair use. *See Kelly*, 336 F.3d at 821 (finding "it was reasonable" to "copy

each of Kelly's images as a whole" as copying only part would "reduc[e] the usefulness of the visual search engine); *HathiTrust*, 755 F.3d at 98 ("[b]ecause it was reasonably necessary for the [HathiTrust Digital Library] to make use of the entirety of the works in order to enable the full-text search function, we do not believe the copying was excessive"); *Sony Betamax*, 464 U.S. at 449–50 (finding technology for recording entire programs was fair use and noting: "the fact that the entire work is reproduced, does not … militat[e] against a finding of fair use"); *Bill Graham Archives*, 448 F.3d at 613; *iParadigms*, 562 F.3d at 639–40 (copying of entire papers fair).

Google Books is highly instructive on factor three. In that case, Google "made a digital copy of the entirety of each of Plaintiffs' books" to create a tool that could search for words within those books and return snippets containing them when prompted. 804 F.3d at 221. With significant effort, plaintiffs' experts were able to use the tool "to access [and reproduce no more than] 16% of the text" of the books. *Id.* at 222–23. The court held that factor three weighed in favor of fair use:

> While Google makes an unauthorized digital copy of the entire book, it does not reveal that digital copy to the public. The copy is made to enable the search functions to reveal limited, important information about the books. With respect to the search function, Google satisfies the third factor test, as illuminated by the Supreme Court in *Campbell*. *Id.* at 221–22.

Like the tools in *Sony Betamax*, *Perfect 10*, *Kelly*, *HathiTrust* and *Google Books*, Llama's utility depends on copying whole books (and many other data sources). Plaintiffs openly acknowledge that. *See* Pls' Mot. at 5 (asserting that "Books Constitute High-Quality Training Data" that the "parties' experts agree" is "*imperative for training* because it builds diverse and nuanced relationships between words, improving output quality."); *see also id.* at 4–5 (asserting that "books are uniquely valuable as data for developing longer-context windows," i.e., "the ability to produce outputs based on long prompts"). Thus, Meta's use of whole books to train Llama was necessary for its fair use purpose of creating a transformational LLM.

Moreover, Meta's use does not make any significant portion of the texts available to Llama users. Experiments by the parties' experts show that Llama can be forced—through "determined, assiduous, time-consuming" prompt engineering, *Google Books*, 804 F.3d at 223—to make accessible, *at most*, only a few, frequently quoted passages (usually less than a paragraph) from only some of Plaintiffs' books. *Supra*, at 10. These passages constitute far less than 1% of any

work, and pale in comparison to the reproduction found fair as a matter of law in *Google Books*. And Llama's inability to generate outputs that substantially replicate training data is *by design*: Meta trained Llama on a massive volume of data to ensure that no one work meaningfully influences what Llama can generate (Ungar ¶¶ 60–64), and Meta's implemented mitigations further guard against the possibility of infringing outputs (*supra*, at 9). Indeed, Plaintiffs do not even allege any infringing outputs. Thus, if the third factor thus weighs in favor of anyone, it is Meta.

4. Llama does not harm any cognizable market for Plaintiffs' works (Factor Four)

The fourth factor examines potential market effects from the challenged use, asking whether that use will cause "harm to the actual or potential markets for the copyrighted work." *Oracle*, 593 U.S. at 24. Consideration of the fourth factor requires courts to identify the relevant markets, assess any "harms" to those markets that are cognizable under copyright law, and then balance those harms against the public interest. *Oracle*, 593 U.S. at 35 ("[W]e must take into account the public benefits the copying will likely produce."). Here, this factor strongly supports a finding of fair use, because Llama does not provide a substitute for Plaintiffs' works in any relevant market and has not caused Plaintiffs any cognizable harm, but does provide tremendous public benefits.

When defining the relevant market, the inquiry is not whether the disputed use harmed *any* market for the original work, but rather whether the use served as a substitute in a market the copyright holder reasonably expected. *Harper & Row*, 471 U.S. at 568 (looking at "use that supplants any part of the *normal* market for the copyrighted work"). As the Supreme Court has acknowledged, "it is a given in every fair use case that plaintiff suffers a loss of a potential market if that potential is defined as the theoretical market for licensing the very use at bar." *Oracle*, 593 U.S. at 38 (warning of the "danger of circularity" posed by framing the inquiry in this way) (citing 4 Nimmer on Copyright § 13.05[A][4]). To avoid this "vice of circular reasoning," courts emphasize that the relevant inquiry is not whether the disputed use harmed *any* market for the original work, but rather whether the use served as a substitute in a market the copyright holder reasonably expected. *Swatch Grp. Mgmt. Svs. Ltd. v. Bloomberg L.P.*, 742 F.3d 17, 34 (2d Cir. 2014) (fourth factor assesses "impact on potential licensing revenues for *traditional, reasonable, or likely to be developed markets*"). In keeping with this authority, the Ninth Circuit has squarely

held that the copyright holder "cannot prevent others from entering fair use markets merely by developing or licensing a market for . . . transformative uses of its own creative work." *Tresona Multimedia, LLC v. Burbank High School Vocal Music Ass'n*, 953 F.3d 638, 652 (9th Cir. 2020) (quoting *Bill Graham Archives*, 448 F.3d at 614–15).

In addition to defining the relevant market, courts must also assess whether the harm asserted is cognizable under copyright law. "[A] potential loss of revenue is not the whole story." *Oracle*, 593 U.S. at 35. For instance, a "lethal parody, like a scathing theatre review," may "kill[] demand for the original," but that would not be "cognizable under the Copyright Act." *Campbell*, 510 U.S. at 591–92. Instead, the fourth factor asks "whether consumers treat a challenged use 'as a market replacement' for a copyrighted work." *Goldsmith*, 598 U.S. at 555 (Gorsuch, J., concurring). That inquiry "is concerned with *only one type of economic injury* to a copyright holder: the harm that results because the *secondary use serves as a substitute for the original work*." *Hathitrust*, 755 F.3d at 99. Plaintiffs cannot show any cognizable substitutive harm from Llama.

a. Llama does not substitute for Plaintiffs' books or harm any normal market for them

The normal or traditional markets for Plaintiffs' trade books include publication and sale of physical books, creation and sale of audio and e-books, and potential optioning or licensing for adaptation (e.g., film, television). Sinkinson ¶¶ 12, 58. As admitted at deposition and in discovery responses, Plaintiffs cannot identify any lost sales or licensing opportunities in these markets. BG Exs. 8, 9. None has lost a book deal because of Llama, and no Plaintiff testified that they intended to stop writing books because of it. *Id.* Ex. 13. Plaintiffs' expert, Dr. Spulber, likewise did not cite any evidence of lost sales. He did not even evaluate it. BG Ex. 25, Spulber Dep. 242:17–243:5 ("So I have not looked at whether plaintiffs had lost sales other than the fact that, when Meta made copies, they lost sales."). Industry data assessed by Dr. Sinkinson, and unrefuted by Plaintiffs' experts, further validates that Llama's broad release had no discernible effect on book sales. Sinkinson ¶¶ 18–35.[14] This all weighs heavily in Meta's favor on factor four. *See Seltzer*, 725 F.3d at 1179

[14] If anything, the fact that Llama can answer questions about Plaintiffs and their books (*supra*, at 15) may *boost* sales. Sinkinson ¶¶ 36–54; *see also Ty, Inc. v. Publ'ns Int'l Ltd.*, 292 F.3d 512, 517 (7th Cir. 2002) ("Book reviews that quote from ('copy') the books … increase the demand for copyrighted works; to deem such copying infringement would therefore be perverse, and so the fair-use doctrine permits such copying.").

(finding fair use, in part, based on admission that "value of [plaintiff's] work was unchanged" and "nobody had ever told him that he would not buy his work as a result of [defendant's] use").

Plaintiffs' inability to demonstrate harm is not surprising: they do not identify any instance in which Llama has ever outputted their books, or anything that could substitute for reading them. Moreover, even if Plaintiffs had evidence that Llama's ability to answer questions about Plaintiffs' books affected sales (they do not), that would not be a cognizable harm. As explained in *Google Books*, the ability to search for factual information from or about a book on a search engine or other tool that returns snippets from the book might in some cases "eliminat[e] any need to purchase it or acquire it from a library," but that "would not change the taking of an unprotected fact into a copyright infringement." 804 F.3d at 224; *see also HathiTrust*, 755 F.3d at 100 (ability to search the text of a book to determine whether it includes selected words "does not serve as a substitute for the books that are being searched").[15] The critical fourth factor also strongly favors Meta.

b. Markets for transformative uses are not relevant to this analysis and, in any event, no such market exists for the use Meta made of Plaintiffs' works

Unable to show harm to normal markets from Meta's copying of their books to train Llama, Plaintiffs resort to asserting that they were deprived of a license fee for that precise use. This circular argument fails for three reasons: (1) Plaintiffs' copyrights do not extend to any market for transformational uses; (2) there was no market to license Plaintiffs' works for LLM training when Meta copied datasets containing them (and there *still* is not); and (3) no such market is likely to develop.

First, copyrights do not extend to monopolizing rights in markets for licensing works for transformative uses, including the use of trade books as LLM training data. As the Supreme Court and Ninth Circuit have repeatedly concluded, the secondary user's failure to obtain a license in order to pursue a transformative use—even where the parties had engaged in negotiations for such a license—does not weigh against fair use. Thus, in *Oracle*, 593 U.S. at 8, the Supreme Court affirmed fair use where Google had sought, but subsequently abandoned, efforts to secure a license for the Java code it used to build its Android platform. Likewise in *Seltzer*, 725 F.3d at 1179, Green

[15] Plaintiffs have also never explained why the extensive information readily available about their books online is qualitatively different from any information that can be accessed using Llama.

Day's transformative use as a concert backdrop of an artist's iconic "Scream" street art icon was fair, even though the artist had licensed it for use in another music video. *See also Campbell*, 510 U.S. at 585 n.18 ("reject[ing] Acuff-Rose's argument that 2 Live Crew's request for permission to use the original should be weighed against a finding of fair use" because "the offer may simply have been made in a good-faith effort to avoid [] litigation").

Courts outside this circuit agree. "[A] copyright holder cannot prevent others from entering fair use markets merely 'by developing or licensing a market for … transformative uses of its own creative work.'" *Bill Graham Archives*, 448 F.3d at 614–15. Even where the copyright holder is already engaging in a transformative use, that does not take away its transformative character when others engage in it, because "copyright owners may not preempt exploitation of transformative markets." *Id.* at 615; *see also Authors Guild, Inc. v. HathiTrust*, 902 F. Supp. 2d 445, 463 (S.D.N.Y. 2012) (citing *Bill Graham Archives* and finding: "A use that 'falls within a transformative market' does not cause the copyright holder to 'suffer market harm due to the loss of license fees.'"). In *Bill Graham Archives*, for instance, the court held, at summary judgment, that unlicensed reproduction of entire concert posters in a book about the Grateful Dead constituted fair use as it was "transformatively different from the [posters'] original expressive purpose." 448 F.3d at 609. It rejected the argument that this invaded any market to which the plaintiff could legitimately lay claim, even though the defendant had "directly contacted [plaintiff] seeking to negotiate a license agreement." *Id.* at 607, 615. That is, the Second Circuit held that plaintiff had not suffered harm to a cognizable market even though it "established a market for licensing its images, and in this case expressed a willingness to license images to" the defendant, because neither act showed "impairment to a traditional, as opposed to a transformative market." *Id.* at 614.

Were the law otherwise, every copyright plaintiff could attempt to defeat fair use by leveraging litigation to obtain "licenses" for similar transformative uses. Imagine an artist who copies copyrighted books to make confetti or paper airplanes, or makes papier mache sculptures of celebrities from their photographs, or rearranges every note from a recording of the Sound of Music into a heavy metal opera— transformative uses that cannot possibly substitute for the originals. If the copyright owner could defeat fair use simply by licensing that exact use to someone else, there

would be no logical end to their copyright monopoly, which protects only their original expression, not any and all uses of their work. *See Connectix*, 203 F.3d at 607 ("because the Virtual Game Station is transformative, and does not merely supplant the PlayStation console, the [VSG] is a legitimate competitor in the market for platforms on which Sony and Sony-licensed games can be played…. Sony understandably seeks control over the market for devices that play games Sony produces or licenses. The copyright law, however, does not confer such a monopoly."). Where, as here, Plaintiffs claim only harm in a market for transformative use, the fourth factor favors fair use.

Second, even if markets for licensing works for transformative uses were legally relevant (they are not), there is no evidence that a market for licensing books to train LLMs exists today, let alone in 2022 when Meta began training Llama. For there to be a market, there must be something of value to exchange, but none of Plaintiffs' works has economic value, individually, as training data. Ungar ¶¶ 60–64; Sinkinson ¶¶ 56–57. The marginal utility of any individual book to the quality of Llama outputs is effectively zero. Ungar ¶¶ 60–61. The contribution of individual works to model performance, if any, is an emergent property that manifests in large, aggregated corpuses of text, rather than the individual works themselves. Ungar ¶¶ 44–47. Thus, the economic value of individual titles is thus, at best, indeterminable and de minimis. Sinkinson ¶¶ 56–57.

Plaintiffs make much of Meta's preliminary discussions with certain publishers in 2023 to ascertain whether they could provide a large corpus of books for training. *See* Pls' Mot. at 8, 16. But they do not claim that Plaintiffs' books were at issue in those negotiations, nor could they: Plaintiffs' publishers lack authority to license their books for AI training. BG Exs. 17, 18. Likewise, Meta's exploratory discussions with textbook publishers (Pls' Mot. at 8–9, 16), and a 2022 contract for translations of books written in lesser spoken African languages (*id*. at 8 n. 17; *see* BG Ex. 31 (Boesenberg Dep.) at 186:1–11, 188:1–15, 189:9–13), do not evidence a viable market for licensing Plaintiffs' trade books as training data. Because Meta—and every other LLM developer—requires more text data than is readily publicly available, procuring *access* to rare or large volumes of well curated scientific or other textbook data might have value even if individual works do not. Sinkinson ¶ 63; *see* BG Ex. 25 (Spulber Dep.) at 142:2–144:1 (acknowledging that parties may pay for access to training data, independent of any IP protection).

In any event, no Plaintiff has ever licensed a book as training data for an LLM or AI model. BG Exs. 10–12. To the contrary, each admits they have neither offered their works for this purpose, nor been asked to do so. *Id*. Nor did any Plaintiff write books with an expectation they might someday monetize them through a license for AI training. *See, e.g.*, BG Ex. 14; *cf. id.* Ex. 13.

Numerous courts have rejected efforts like Plaintiffs' here to claim harm to a potential market they have never sought to avail themselves of. In *Blanch*, for instance, the Second Circuit held that the fourth factor favored fair use concerning a visual artist's use of copyrighted magazine ad photographs in collage-style mixed media artwork, because the photographer had never licensed her photographs for incorporation in visual artworks and, thus, the use "did not cause any harm to her career or upset any plans she had for [the photograph]." 467 F.3d at 258.

The Seventh Circuit also affirmed summary judgment in *Kienitz v. Sconnie Nation LLC*, 766 F.3d 756, 760 (7th Cir. 2014), a case involving use of a copyrighted photograph of Einstein on novelty t-shirts. Judge Easterbrook focused on factor four and held it favored fair use because a "t-shirt … is no substitute for the original photograph. Nor does Kienitz say that defendants disrupted a plan to license this work for apparel. Kienitz does not argue that defendants' products have reduced the demand for the original work or any use of it that he is contemplating." *Id.* at 759.

Summary judgment was also granted in *Video-Cinema Films, Inc. v. Cable News Network, Inc.*, which challenged use of film footage in obituaries broadcast by major news channels after an actor's death. 2001 WL 1518264, at *9 (S.D.N.Y. Nov. 28, 2001). The court held this was fair use and rejected the plaintiff's circular argument that this deprived it of a market for licensing film clips for obituaries, explaining that "consideration of licensing revenues is not permitted absent evidence that a regular traditional market exists for the specific use at issue." *Id.* Since only one defendant had ever paid to license a clip for an obituary, and plaintiff had only ever received three "payments from other stations (*to avoid litigation*)," the court held this was "not a regular traditional market" and any purported harm to plaintiff was outweighed by the public benefit of allowing the use. *Id*.

Goldsmith and *Harper & Row* also prove the point. *Goldsmith* involved use of an Andy Warhol silkscreen made from Goldsmith's photograph of Prince "as a commercial substitute for her own protected photograph in sales to magazines looking for images of Prince to accompany

articles about the musician." 598 U.S. at 558 (Gorsuch, J., concurring). There was no dispute as to an existing market or whether Goldsmith had availed herself of that market; she had previously licensed the photograph to other national magazines for that *exact use*. *Id.* at 517. Likewise, in *Harper & Row*, The Nation "effectively arrogated to itself the right of first publication" by using extensive verbatim excerpts of Ford's unpublished memoir as "featured episodes in a story about the Nixon pardon—*precisely the use petitioners had licensed to Time.*" 471 U.S. at 549, 568. Plaintiffs cannot make any such showing here.[16]

Third, undisputed evidence establishes that a potential market for licensing Plaintiffs' books as LLM training data is not likely to develop.

One "goal" of fair use "is to facilitate a class of uses that would not be possible if users always had to negotiate with copyright proprietors." *Kienitz*, 766 F.3d at 759. "Many copyright owners would block all parodies, for example, and the administrative costs of finding and obtaining consent from copyright holders would frustrate many academic uses." *Id.* Likewise, where it would be "prohibitively expensive to develop a market to license the use of works" for a particular purpose, such that it would "prohibit the formation of a viable market *in the first place*," this cuts against a finding of cognizable market harm and in favor of fair use on factor four. *Author's Guild v. HathiTrust*, 902 F. Supp. 2d 445, 463–64 (S.D.N.Y. 2023).

Here, any theoretical market for licensing text as training data is doomed to "market failure," an economic phenomenon that occurs when, notwithstanding the existence of willing buyers and sellers for a good, no transactions occur, or the quantity of goods transacted is socially inefficient. Sinkinson ¶ 55.[17] The conditions leading to market failure here are manifold.

First, and as explained above, individual books have negligible value as training data, *id.* ¶¶

[16] Plaintiffs also cannot claim harm to a market from which they would abstain. *See, e.g., Mattel, Inc. v. Walking Mtn. Prods.*, 353 F.3d 792, 806 (9th Cir. 2003) (affirming summary judgment to artist who made adult-themed art with Barbie Dolls, a market it found "safe to assume that Mattel will not enter … or license"); *iParadigms*, 562 F.3d at 644 (finding fair use, in part, as plaintiffs would not allow the challenged use); *Wilder v. Hoiland*, 2024 WL 382141, at *10 (S.D.N.Y. Feb. 1, 2024) (same; granting summary judgment). Here, several Plaintiffs testified they find AI objectionable or that they would not want their work used to train an LLM. BG Exs. 15, 16. This further cuts against them on factor four. *See Campbell*, 510 U.S. at 592 (relevant market "includes only those that creators of original works would in general develop or license").

[17] A market is a set of *similar* transactions between buyers and sellers; thus, the relevant theoretical market is the licensing of general audience trade books as training data. Sinkinson ¶ 55.

56–57, and even the value of large corpuses is uncertain. *Id*. ¶¶ 64–65. Meta has invested hundreds of millions of dollars in LLM development, but does not project a return on that investment for years, with the degree of any financial benefit being highly speculative. BG Ex. 26 (Acharya Dep.) at 49:3–51:8; 338:4–340:5; Ex. 27. Moreover, technological development within the field is advancing rapidly, and there is significant uncertainty about the optimal proportions of different categories of text. *Id.* Ex. 31 (Bell Dep.) at 41:4–43:12, 49:7–53:12; Ex. 32 (Boesenberg Dep. 50:13–51:4 ("we [don't] have a really strong view on which [texts are] more important than others"); Ex. 46 (Nho Dep.) at 118:1–16. At the same time, the economics of the publishing industry are such that transaction costs are high. Sinkinson ¶¶ 58–61. Although five publishing houses collectively account for roughly 80% of the U.S. market for trade books, any right to license those books to train an LLM, if it exists, is not delineated in publishing contracts and is reserved to authors. *Id*. ¶¶ 58–59. Plaintiffs acknowledged this in deposition and discovery. BG Exs. 17, 18. Thus, a party seeking to license a large volume of trade books cannot obtain them from publishers, as Meta discovered when it investigated doing just that in early 2023. BG Exs. 33, 34 (Choudhury Dep.) at 23:5–13 ("[W]e as a company learned that most if not all of the fiction publishers did not have … worldwide rights to license the copyrights," and "we weren't getting a lot of engagement from the publishers in general."); Ex. 31 (Boesenberg Dep.) at 55:1–58:56, 152:21–153:13; Sinkinson ¶ 67. Moreover, unlike the music industry, which supports collective rights organizations that administer copyright licenses on behalf of millions of artists (e.g., ASCAP and BMI), no centralized licensing regimes exist for trade books. Sinkinson ¶ 62.

Thus, there is no economically feasible mechanism for Meta or other LLM developers to obtain licensed copies of the astonishingly large volume of books and other training data necessary for the technology to exist and advance. *Id*. ¶¶ 55–71; Ungar ¶ 43; BG Ex. 30 (Bell Dep.) at 131:21–133:2, 136:6–136:20; Ex. 45 (Nho Dep.) at 148:8–149:2. Absent fair use, Meta would have to initiate individualized negotiations with *millions* of authors. Sinkinson ¶¶ 60-61, 69-70. Among other things, this would entail identifying individual books and their authors; determining how to contact them; ascertaining whether they own rights clear of encumbrances (i.e., assignments or exclusive licenses); negotiating an acceptable price with each author or agent (since there is no

mechanical royalty mechanism for books, this would widely vary depending on the author and how well the book is selling—factors that are largely irrelevant to AI training); and negotiating scope and other license terms, including suitable representations and warranties. *Id*. This process would be onerous for even a few authors; it is practically impossible for hundreds of thousands or millions. *Id*.[18] At the same time, because the value of any work to LLM training is negligible (and indeterminable) (Ungar ¶¶ 60–64), it is economically irrational to engage in protracted and expensive negotiations to license individual works. Sinkinson ¶¶ 55–70. It is, thus, no surprise that despite widespread use of Books3 and other books datasets to train AI models, and the spate of class action lawsuits that have followed, no market has developed to license books—and in particular, Plaintiffs' books—for AI training. BG Exs. 10–12. Rather than showing harm to an actual or likely market, Plaintiffs are trying to *create* a market under the threat of litigation over a use that, by all factors and measures, is quintessentially transformative and fair.

c. Llama's benefit to the public greatly outweighs any hypothetical harm to the market for Plaintiffs' books

It bears repeating that the "ultimate goal of copyright is to expand public knowledge and understanding[.]" *Google Books*, 804 F.3d at 212. That is literally what Llama is designed to do, and the innumerable (undisputed) public benefits of this transformational new technology greatly outweigh any speculative harm to Plaintiffs. As a result of Meta's open-source approach, start-ups, non-profits, and research labs—smaller entities that would otherwise lack the resources to develop LLMs—have the opportunity to experiment with and adapt Llama as their own. Nayak ¶¶ 24–32. Millions are using Llama, or platforms built on Llama, to bring innovative and, in some cases, potentially life-saving services and technologies to market that have nothing whatsoever to do with Plaintiffs' books. *Id.* ¶¶ 28, 32. Llama also levels the playing field for the disabled and those who need translation or writing help to better communicate. *Id.* ¶ 32. These public benefits "must [be] take[n] into account," and strongly weigh in favor of fair use here. *Oracle*, 593 U.S. at 35.

[18] Importantly, as Professor Sinkinson explains, "all of the factors that undermine the viability of a market that includes licensing books to train LLMs are exacerbated for other categories of text training data." For example, Common Crawl, which was by far the largest source of pretraining text for Llama, contains roughly 15 trillion tokens of deduplicated data composed of billions of discrete works. Determining which portions may be copyright protected, let alone who owns those rights, and negotiating and executing agreements with those rightsholders, is "impossible." *Id*. ¶ 69.

Balancing all factors together, Meta's use is fair because it provides vast public benefits, increases public knowledge, encourages the creation of new expression, and otherwise serves copyright's core purpose of promoting the progress of science and the useful arts, all without any cognizable harm to Plaintiffs. No court has ever held such a balance of benefits and harms to be anything other than fair use. The Court should therefore find that Meta's copying of Plaintiffs' books to train Llama was fair use under 17 U.S.C. § 107, and grant summary judgment for Meta on Plaintiffs' claim for violation of their § 106(1) reproduction rights.

D. Plaintiffs' Motion Should Be Denied in Full

Plaintiffs' Motion asserts two grounds for summary judgment in Plaintiffs' favor; both fail. First, Plaintiffs seek a finding that Meta violated their § 106(1) reproduction rights based on "Meta's large-scale copying of the Books without Plaintiffs' permission." Pls' Mot. at 2, 19. As detailed above, Meta's copying of Plaintiffs' books constituted a transformative fair use and, thus, was "not an infringement of copyright" under § 107. *See Sony Betamax*, 464 U.S. at 447. Plaintiffs' request for summary judgment on its § 106(1) claim should, therefore, be denied.

Plaintiffs' second argument for summary judgment, though convoluted, is equally meritless. They assert that Meta's "initial acquisition" of their books "cannot be fair use" as Meta downloaded them from "known pirated databases," including via torrent, and "reuploaded to other online pirates at least some quantity of that pirated data as part of the peer-to-peer ('P2P') sharing process." Pls' Mot. at 2; *id.* at 22 (asserting that Meta's "initial reproduction" was not fair because it also involved "distributing copyrighted material to unknown third parties." None of this holds water.

Section 106 does not provide a right of "acquisition." It provides for exclusive rights of reproduction (i.e., the right to "reproduce the copyrighted work in copies") (17 U.S.C. § 106(1)), and distribution (i.e., the right to "distribute copies … of the copyrighted work to the public") (*id.* § 106(3)). Plaintiffs conflate these rights in seeking a determination that Meta's "acquisition" of Plaintiffs' works through torrenting was not fair use, but they are distinct rights and must be analyzed separately. *See A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1013–14 (9th Cir. 2001), *as amended* (Apr. 3, 2001) (absent fair use, "users who upload file[s] … for others to copy violate plaintiffs' distribution rights," while "users who download files containing copyrighted music

violate plaintiffs' reproduction rights"), *aff'd sub nom.*, 284 F.3d 1091 (9th Cir. 2002); *Columbia Pictures Indus., Inc. v. Fung*, 710 F.3d 1020, 1034 (9th Cir. 2013) ("Both uploading and downloading copyrighted material are infringing acts. The former violates the copyright holder's right to distribution, the latter the right to reproduction.").

Distribution is disputed: Any argument premised on Meta's alleged distribution of Plaintiffs' works is not properly before the Court. Plaintiffs do not move for summary judgment that Meta violated § 106(3), and for good reason. Plaintiffs have not asserted, much less put forth undisputed facts to prove, that Meta distributed any of their works. Indeed, the most they can muster is that "at least some quantity" of the data Meta downloaded via torrent must have been "reuploaded." Pls' Mot. at 2. That does not even suffice to raise a genuine issue as to whether any distribution of *Plaintiffs'* works occurred, much less allow the Court to determine as a matter of law that Meta's copying of Plaintiffs' works was anything but fair.

Plaintiffs' distribution claim is premised on two forms of "data uploading": "leeching," which allegedly "involves simultaneous 'tit-for-tat' reuploading that occurs during downloading" and "seeding," which "occurs after a user completes downloading a data file but continues to offer or 'seed' the data file to other users." Pls' Mot. at 12. Seeding is the only form of distribution alleged in the TAC, and until the Court's recent order allowing the expert report of Dr. Choffnes and further discovery on "leeching" (Dkt. 470), the only theory at issue in the case. Thus, fact discovery is newly underway as to leeching, and Meta has sought leave to file its own expert report on leeching—a topic it has not yet had an opportunity to address (Dkt. 486). As to seeding, Plaintiffs no longer appear to be pursuing this theory. Plaintiffs now concede that Meta used "a script that worked to prevent 'seeding' the pirated data after downloading was complete." Pls' Mot. at 13 (citing Pritt Ex. 71 (Choffnes Report) ¶¶ 16, 19).

Plaintiffs' Motion also comes nowhere close to establishing that Meta distributed any of their works by leeching. The only purported evidence they cite for this is the Choffnes Report. *Id.* at 25, 26 (citing Pritt Ex. 71 ¶¶ 20–30). That unsworn report is not attached to a supporting declaration, is not admissible evidence, and should be excluded from consideration of Plaintiffs'

Motion.[19] Even if the Court were to consider it, Dr. Choffnes opines only that there is "a greater than 99.99999% chance that Meta uploaded *at least one piece* of Plaintiffs' works." *Id.* at 25. That opinion is based on faulty reasoning, as Meta's expert, Ms. Frederickson-Cross, attests in her declaration (Fredericksen ¶¶ 17-45), raising at the very least a genuine issue of disputed fact on distribution. And even if the Choffnes Report were credited, the most any factfinder could infer from it is that Meta uploaded an unspecified amount of an unidentified work to an unidentified recipient, which falls far short of establishing that Meta distributed any Plaintiff's work via leeching.

In sum, Plaintiffs cannot rely on disputed allegations that Meta distributed their works—a determination on which they do not seek summary judgment and implicitly concede they cannot establish as a matter of law—to overcome Meta's defense that its copying of Plaintiffs' works to train Llama was fair. Meta looks forward to addressing the facts and law undercutting the viability of any distribution claim when the newly ordered discovery and expert work is completed.[20]

Meta's use was fair irrespective of its method of acquisition: Plaintiffs' argument that Meta's "initial acquisition" of works from "pirate" websites, including via torrent, cannot be fair as a matter of law must also be rejected. As explained above in the discussion of the first fair use factor, any evidence of "bad faith" in copying works for a transformative purpose is of little, or no, consequence to the fair use analysis. *Supra* at 15–21. Further, Meta downloaded copies of datasets that included Plaintiffs' books for the fair use purpose of training Llama, which does not contain, replicate, reproduce, or distribute those works or let anyone see or read them. *Supra* § II.D. Meta's copying was thus for a use readily distinguishable from that at issue in the trove of music file-sharing cases Plaintiffs cite dating back to the days of Napster.[21] In all of those cases, the accused

[19] *See Progressive Sols., Inc. v. Stanley*, 2018 WL 1989547, at *8 (N.D. Cal. Mar. 8, 2018) (granting motion to exclude and stating: "Unsworn expert reports prepared in compliance with Rule 26(a)(2) do not qualify as affidavits or otherwise admissible evidence for purpose of Rule 56, and may be disregarded by the court when ruling on a motion for summary judgment"); *Harris v. Extendicare Homes, Inc.*, 829 F. Supp. 2d 1023, 1027 (W.D. Wash. 2011) (noting courts in the Ninth Circuit "routinely held that unsworn expert reports are inadmissible") (citing cases)).

[20] On March 21, 2025, Meta filed its answer to the TAC, which includes the new § 106(3) claim. (Dkt. 485.) Meta reserves all arguments and defenses as to that claim, including the right to address the related case law and arguments improperly raised in Plaintiffs' Motion.

[21] *See* Pls' Mot. at 24–25 (citing *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 919 (2005); *Napster, Inc.*, 239 F.3d 1004, 1014–17; *BMG Music v. Gonzalez*, 430 F.3d 888, 890 (7th Cir. 2005); *Sony BMG Music Entm't v. Tenenbaum*, 672 F. Supp. 2d 217, 227 (D. Mass.

platform allowed users to upload and download entire copyrighted music files to play and listen to, a non-transformative use which supplanted the normal market for the songs. Llama does not do that.

Plaintiffs' remaining authorities are equally unavailing here as they deal with situations in which works were copied for purposes of making them available to read or watch, rather than for a transformative use. As Plaintiffs' Motion acknowledges (at 28), the Second Circuit affirmed a finding of no fair use in *Hachette Book Group, Inc. v. Internet Archive*, because defendant "copie[d] the Works in full and ma[de] those copies available to the public in their entirety," which did not "achieve a transformative secondary purpose," but "supplant[ed] the originals." 115 F.4th 163, 190 (2d Cir. 2024). In *American Geophysical Union v. Texaco Inc.*, defendants simply photocopied entire articles and sent them around to other researchers to read—a non-transformative use that served "the same basic purpose that one would normally seek to obtain the original[.]" 60 F.3d 913, 918, 923 (2d Cir. 1994). In *Glacier Films (USA), Inc. v. Turchin*, the defendant "downloaded and distributed repeatedly … a Hollywood action movie." 896 F.3d 1033, 1035 (9th Cir. 2018). In *Ross Intelligence*, the district court rejected defendant's fair use defense based on a finding that it copied plaintiff's copyrighted Westlaw headnotes for an identical, non-transformative purpose—to create a competing legal research tool that would perform the same functions as, and substitute for, Westlaw. 2025 WL 458520, at *7–10.[22] And Plaintiffs once again cite this Court's decision in *In re DMCA* out of context for the proposition that "it is obvious … that downloading and distributing copyrighted music via peer-to-peer systems does not constitute fair use." 608 F. Supp. 3d at 879.

In contrast here, there is no allegation or evidence that the copies Meta made were used for reading Plaintiffs' books—by Meta employees or anyone else. Nor can users read Plaintiffs' books on Llama. Meta used the copies it made to develop and train Llama—a transformative use that is not remotely analogous to those at issue in Plaintiffs' cases. Indeed, that is why, instead of arguing that Meta's *use* was unfair (which they cannot show), Plaintiffs focus on "acquisition."

What Plaintiffs ask the Court to do is ignore the *use* Meta made of the copies of Plaintiffs' works, and find that it "acquired" those copies in an unfair manner. But Plaintiffs do not cite any

2009); *In re Aimster Copyright Litig.*, 334 F.3d 643, 645, 655 (7th Cir. 2003); *United States v. Slater*, 348 F.3d 666, 669 (7th Cir. 2003)).

[22] *Ross* has moved for interlocutory appeal. No. 1:20-cv-00613 (D. Del. Mar. 18, 2025), Dkt. 786.

case, and we are aware of none, in which a court made a determination as to whether the copying of a work was fair or not without considering *why* the defendant made the copies, *what use it made of them*, and how this use affected the plaintiff. By statute, fair use requires evaluation of the "purpose and character of the *use*" (factor one), and the "effect of the *use*" upon the market for the original (factor four). 17 U.S.C. § 107; *see Goldsmith*, 598 U.S. at 510 ("[A] use that has a distinct purpose is justified because it furthers the goal of copyright…."); *Sony Betamax*, 464 U.S. at 433 ("Any individual may reproduce a copyrighted work for a 'fair use.'"). Plaintiffs' request for summary judgment that Meta's "initial acquisition" was unfair, disembodied from any evaluation of its quintessentially and undeniably transformative *use* to develop and train Llama, misconstrues the basic tenets of copyright law and fair use jurisprudence and should be denied in full.

E. The Court Should Grant Summary Judgment to Meta on the DMCA Claim

Following the Court's March 7 Order (Dkt. 471), Plaintiffs' Section 1202(b) claim has been reduced to a single untenable theory: that Meta intentionally removed copyright management information ("CMI") from datasets used to train the Llama models, knowing or having reason to know that doing so would conceal its infringement of Plaintiffs' books. *See* 17 U.S.C. § 1202(b). This theory must now reckon with the facts. No reasonable jury could find that Meta violated Section 1202(b) for three independent reasons: (1) unrebutted testimony from Meta employees and the parties' experts conclusively establish that Meta removed CMI from training data alongside other repetitive text as a part of industry standard procedures to improve performance; (2) the record is devoid of any evidence (fact, expert, or otherwise) or coherent explanation of how CMI removal conceals infringement; and (3) because nothing was concealed from Plaintiffs, they suffered no harm from CMI removal and thus lack statutory and Article III standing.[23]

Although Meta bears the initial burden of showing the absence of a genuine issue for trial, it need not disprove Plaintiffs' case. *Anderson v. Liberty Lobby, Inc*., 477 U.S. 242, 256 (1986) (once movant meets burden to show lack of a genuine issue of fact, plaintiff must in turn proffer "evidence that would support a jury verdict"). When opposing summary judgment, the plaintiff

[23] If the Court finds that Meta's use of Plaintiffs at-issue books is fair use, then Plaintiffs' Section 1202(b) claim fails for the additional reason that there was no "infringement" to conceal.

"must offer more than conclusory allegations, and if the defendant presents affidavits or other evidence establishing a lack of scienter, the plaintiff must come forward with some affirmative showing." *Vucinich v. Paine, Webber, Jackson & Curtis, Inc.*, 739 F.2d 1434, 1436 (9th Cir.1984). Where the record is devoid of evidence that CMI was removed with culpable scienter, courts routinely grant summary judgment on § 1202 claims. *See Powers v. Caroline's Treasures Inc.*, 382 F. Supp. 3d 898, 904 (D. Ariz. 2019) (testimony and corroborating documents sufficed to defeat § 1202 claim on summary judgment); *Victor Elias Photography, LLC v. Ice Portal, Inc.*, 43 F.4th 1313, 1323 (11th Cir. 2022) (affirming summary judgment for defendant on § 1202 claim); *Photographic Illustrators Corp. v. Orgill, Inc.*, 118 F. Supp. 3d 398, 407 (D. Mass. 2015) (granting summary judgment on § 1202 claim where allegations of intent were implausible), *aff'd*, 953 F.3d 56 (1st Cir. 2020); *Gordon v. Nextel Commc'ns & Mullen Advert., Inc.*, 345 F.3d 922, 927 (6th Cir. 2003) (affirming dismissal of § 1202 claim as "record contain[ed] no evidence to counter [defendant's] testimony"). The same result is required here.

No Evidence of Scienter. Plaintiffs rely on innuendo, not evidence, that Meta removed CMI with culpable scienter. The record, however, shows that CMI removal had nothing to do with "concealing infringement." The Meta engineer whose team wrote the script to remove certain text from Libgen testified that he chose the sequences of text that were removed because they "commonly occurred in the books" and do not "bring any value to training." BG Ex. 29 (Bashlykov Dep.) at 45:8–49:20, 156:15–158:15. He elaborated that when you filter "you look for particular patterns in the documents," to facilitate removal of useless tokens. *Id*. He had no reason to believe that such removal could "conceal" training data, and had no such intent. Bashkykov ¶¶ 11–12.

Other Meta witnesses testified that removal of duplicative text in training data is standard to avoid overfitting (i.e., memorization) and improve model performance. BG Ex. 37 (Clark Dep.) at 43:15–45:13 ("without doing that cleanup from a data parsing perspective, that would equal poor performance in the model... So it removes noise and it removes repeatability"); Ex. 41 (Esiobu Dep.) 71:3–72 :19; Ex. 42 (Kambadur Dep.) at 70:3–72:6 ("It could make the training less efficient to repeatedly … see the same data."). Plaintiffs' expert concurred. BG Ex. 23 (Lopes Dep.) at 23:21–24 ("Q. Is deduplication of training data an example of a technique that helps prevent

overfitting? A. Yes."). Meta's expert testified this is widespread industry practice, which is unrebutted. BG Ex. 49 (Ungar Dep.) at 190:19–191:2 ("Pretty much nobody trains stuff off the shelf because there's too much boilerplate background and headings. So we're always stripping [that] out….")); Ungar ¶¶ 65–69 (explaining removal).

Any purported intent to conceal is also belied by Meta's public disclosure of its use of Books3 upon release of Llama 1, alleged in Plaintiffs' initial complaint (Dkt. 1; TAC ¶ 39), and the fact that any datasets used to train more recent models were disclosed to Plaintiffs in discovery. TAC ¶¶ 84, 87–88; Dkt. 267 at 30–32 (explaining the timeline of Meta's production of books-related datasets to Plaintiffs). Plaintiffs' theory that CMI removal supposedly stymied them from discovering Meta's alleged infringement of their books is nonsensical and unsupported.

Plaintiffs offer no documents or testimony to the contrary. Their lone expert to opine on the fact that CMI was removed, Dr. Krein, offered no opinion about *why* CMI removal occurred. BG Ex. 21 (Krein Dep.) at 77:13–83:24. Instead, he acknowledged that scripts he identified as being used for CMI removal were designed to also remove numerous sequences of text that have nothing to do with CMI. These scripts also targeted, among other things, chapter numbers, the words "Facebook" or "notebook," the "@" symbol, "www." and lines that start with "Prologue," "_Cover_," "Preface," "Epilogue," "thank you for downloading," "the end," and "leave a review." BG Ex. 22 (Krein Report) ¶¶ 95–101. None of these is CMI. *See* 17 U.S.C. § 1202(c).

No Concealment. The record is also devoid of any evidence that, as a technical matter, CMI removal could have concealed Meta's alleged infringement from them; Plaintiffs have no testimony, documents, or expert opinion to that effect. In *Stevens v. Corelogic, Inc.*, Ninth Circuit affirmed summary judgment on a Section 1202(b) claim due to a similarly deficient record. 899 F.3d 666, 675 (9th Cir. 2018). There, plaintiffs "had not offered *any* specific evidence that removal of CMI metadata from their [works] will impair their policing of infringement." *Id.* To the contrary, the evidence "cut[] against any inference" that material removed from the plaintiff's work was of "any practical significance" to policing against infringement. *Id*. The court noted that the plaintiffs "have not, for example, averred that they have ever used CMI metadata to prevent or detect copyright infringement, much less how they would do so." *Id*. Here too, Plaintiffs have nothing

but innuendo to support their theory that CMI removal could have prevented them from discovering Meta's alleged infringement (particularly where, in reality, *it did not*). The flipside of their argument is also unsupported: that ***but for*** the filtering of lines of training data text containing terms like "copyright" or "ISBN," a Llama user would be able to reliably query the model to disclose that it was trained on Plaintiffs' books. Neither side's experts – and none of the witnesses – support this contrived theory. Thus, Plaintiffs cannot prove the most basic premise of their claim—that removal of repetitive text, including CMI, concealed infringement, much less intentionally.

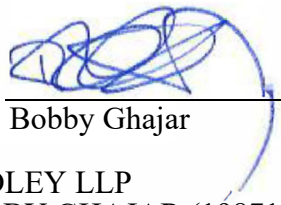
No Concrete Injury. Finally, for these reasons—and notwithstanding the Court's conclusion that Section 1202(b) protects an interest that is "closely related to the kind of property-based harms traditionally actionable in copyright" (Dkt. 471 at 1)—Plaintiffs also lack statutory and Article III standing because they cannot show a concrete injury to *that* interest. *TransUnion LLC v. Ramirez*, 594 U.S. 413, 426 (2021); *see* 17 U.S.C. § 1203(a) (only persons "injured by a violation of section 1201 or 1202" are authorized to "bring a civil action"). In *Intercept Media, Inc. v. OpenAI, Inc.*, Judge Rakoff described Section 1202 injury as something like "[t]he increased possibility of infringement." Order at 16, No. 24-cv-1515 (S.D.N.Y. Feb. 20, 2025), Dkt. 127. With respect to alleged "concealment" of infringement due to CMI removal, Plaintiffs might characterize the Section 1202 harm as difficulty or inability to detect it. But they disclosed no such injury during fact or expert discovery. BG Ex. 53 (Pls' Supp. Initial Disclosures). And it is undisputed that Plaintiffs have long been aware of the very fact they now claim CMI removal concealed—that Meta used their books to train Llama. Thus, there is no evidence that Meta's CMI removal caused any concealment-based harm to Plaintiffs, and it is not plausible that it could have.

IV. CONCLUSION

Meta's copying of Plaintiffs' copyrighted works to train Llama constitutes fair use under 17 U.S.C. § 107, and there is no evidence to support a finding that Meta violated 17 U.S.C. § 1202. Accordingly, Meta respectfully asks the Court to deny Plaintiffs' motion for summary judgment, and grant summary judgment for Meta on Plaintiffs' claims under 17 U.S.C. §§ 106(1) and 1202.

Dated: March 24, 2025

COOLEY LLP

By: 
Bobby Ghajar

COOLEY LLP
BOBBY GHAJAR (198719)
(bghajar@cooley.com)
TERESA MICHAUD (296329)
(tmichaud@cooley.com)
COLETTE GHAZARIAN (322235)
(cghazarian@cooley.com)
1333 2nd Street, Suite 400
Santa Monica, CA 90401
Telephone: (310) 883-6400

COOLEY LLP
MARK WEINSTEIN (193043)
(mweinstein@cooley.com)
KATHLEEN HARTNETT (314267)
(khartnett@cooley.com)
MATTHEW BRIGHAM (191428)
(mbrigham@cooley.com)
JUDD LAUTER (290945)
(jlauter@cooley.com)
ELIZABETH L. STAMESHKIN (260865)
(lstameshkin@cooley.com)
JUAN PABLO GONZÁLEZ (334470)
(jgonzalez@cooley.com)
3175 Hanover Street
Palo Alto, CA 94304
Telephone: (650) 843-5000

COOLEY LLP
PHILLIP MORTON *(pro hac vice)*
(pmorton@cooley.com)
COLE A. POPPELL *(pro hac vice)*
(cpoppell@cooley.com)
1299 Pennsylvania Avenue, NW, Suite 700
Washington, DC 20004-2400
Telephone: (202) 842-7800

CLEARY GOTTLIEB STEEN & HAMILTON LLP

By: 
Angela Dunning

CLEARY GOTTLIEB STEEN & HAMILTON LLP
ANGELA L. DUNNING (212047)
(adunning@cgsh.com)
1841 Page Mill Road, Suite 250
Palo Alto, CA 94304
Telephone: (650) 815-4131

PAUL, WEISS, RIFKIN, WHARTON & GARRISON LLP
KANNON K. SHANMUGAM
(*pro hac vice*)
(kshanmugam@paulweiss.com)
WILLIAM T. MARKS (*pro hac vice*)
(wmarks@paulweiss.com)
2001 K Street, NW
Washington, DC 20006
Telephone: (202) 223-7300

PAUL, WEISS, RIFKIN, WHARTON & GARRISON LLP
ANNA STAPLETON (337686)
535 Mission Street, 25th Floor
San Francisco, CA 94105
Telephone: (415) 432-5000