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, California 90401
Telephone: (310) 883-6400

MARK WEINSTEIN (193043)
(mweinstein@cooley.com)
KATHLEEN HARTNETT (314267)
(khartnett@cooley.com)
JUDD LAUTER (290945)
(jlauter@cooley.com)
ELIZABETH L. STAMESHKIN (260865)
(lstameshkin@cooley.com)
3175 Hanover Street
Palo Alto, CA 94304-1130
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

[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 NOTICE OF MOTION AND MOTION TO DISMISS PLAINTIFFS' THIRD AMENDED COMPLAINT (RULES 12(B)(1) AND 12(B)(6))

Date: February 27, 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

TABLE OF AUTHORITIES
(continued)

Page(s)

TABLE OF AUTHORITIES
(continued)

Page(s)

Statutes

Other Authorities

PLEASE TAKE NOTICE THAT, on February 27, 2025 at 10:00 a.m., Defendant Meta Platforms, Inc. ("Meta") will and hereby does move, under Fed. R. Civ. P. 12(b)(6), to dismiss Claims II and III of the Third Amended Complaint (Dkt. 407) ("TAC") with prejudice.

I. INTRODUCTION

At the crux of this case is an issue of extraordinary importance to the future of generative AI development in the United States: whether Meta's use of publicly available datasets to train its open source large language models ("LLMs")—transformational technology powering innovation, productivity, and creativity—constitutes fair use under U.S. copyright law. The case should remain so focused, despite Plaintiffs' attempts to distract from that core issue by loading up their pleading with unsubstantiated new claims. For reasons previewed in Meta's opposition to Plaintiffs' Motion to Amend and further explained below, Plaintiffs' last-minute claims (II and III) fail.

First, Plaintiffs allege that Meta violated the Digital Millennium Copyright Act, 17 U.S.C. § 1202(b)(1) ("DMCA") (Claim II), by removing copyright management information ("CMI") from their books before using them as training data, and that Meta did so to induce, facilitate, or conceal its own infringement and that of others. As a threshold defect, Plaintiffs fail to allege that they were injured by Meta's removal of CMI, dooming their claim for lack of Article III standing and statutory standing under Section 1203. The TAC also lacks allegations supporting a plausible inference that Meta's removal of CMI aided or concealed infringement, much less that Meta intended this result. Plaintiffs' own allegations underscore that any CMI removal was done to improve training by eliminating repetitive text, ***not*** to promote infringement of Plaintiffs' books.

Second, Plaintiffs' claim under the California Comprehensive Computer Data Access and Fraud Act ("CDAFA"), California Penal Code § 502(c)(2) (Claim III), is merely a claim of copyright infringement by another name. They frame the claim in terms of Meta's unauthorized "access" to their copyrighted works, but "access" here just means "copy," *i.e.*, the exclusive purview of the Copyright Act. Even setting aside preemption, CDAFA prohibits *unauthorized* intrusion into *plaintiff's* computer systems or data; it was not designed to entitle any "data" owner to sue when a party makes *authorized* access to or use of *third-party* datasets on *third-party* systems.

Finally, Plaintiffs' new theory of copyright infringement based on Meta's alleged

"distribution" of datasets is also facially defective, as the TAC contains no allegation, much less facts plausibly establishing, that Meta distributed through "seeding" any of Plaintiff's works at issue. Plaintiffs do not plead a single instance in which any part of any book was, in fact, downloaded by a third party from Meta via torrent, much less that *Plaintiffs' books* were somehow distributed by Meta. However, in light of the Court's recent order allowing limited discovery and an expert report from Meta on seeding, Meta does not seek dismissal of that claim here. Meta looks forward to setting the record straight and debunking this meritless allegation on summary judgment.

The Court should dismiss new Claims II and III in their entirety.[1]

II. SUMMARY OF RELEVANT FACTS AND ALLEGATIONS

A. The Original Complaint and Motion to Dismiss

Plaintiffs' original Complaint alleged that Meta violated Section 1202(b)(1) of the DMCA by "intentionally remov[ing] CMI from the Plaintiffs' Infringed Works" with knowledge or reason to know that doing so "would facilitate copyright infringement by concealing the fact that every output from the LLaMA language models is an infringing derivative work, synthesized entirely from expressive information found in the training data." Dkt. 1 ¶¶ 49, 52. On September 18, 2023, Meta moved to dismiss the Complaint in its entirety, except for Plaintiffs' claim for direct copyright infringement related to training. As to Plaintiffs' Section 1202(b) claim, Meta argued, and the Court agreed, that the claim was deficient for several reasons, including that Plaintiffs failed to identify any examples of allegedly infringing outputs. Dkt. 56 at 3 (dismissing claim). At the hearing, the Court noted that "maybe" Plaintiffs could state a DMCA claim by alleging that an AI model ***reproduced*** their works without the CMI included. *See* 11/9/23 Tr. at 21 (Dkt. 52); *id.* at 23 (stating that the DMCA theory "could ... make sense" if "a user made a query for Sarah Silverman's book, and Sarah Silverman's book was reproduced without the CMI"). Plaintiffs did not allege that theory (*id.* at 26), however, and have not since.

Plaintiffs' original complaint also alleged state law claims for negligence and unfair

[1] Although neither Plaintiffs nor the Court addressed the revised class definition in the proposed TAC, Meta believes that the definition is inappropriate and prejudicial given that it was modified and expanded after the close of discovery. However, based on the Court's prior comments and bifurcation of class certification, Meta does not herein move to strike the claim. Meta asks the Court for leave to file a motion to strike should any of Plaintiffs' claims survive summary judgment.

competition predicated on Meta's use of their books to train Meta's LLMs—the same conduct Plaintiffs had separately pleaded as "copyright infringement." As the Court held, each was styled in the language of the state causes of action, but ultimately were "premised on the rights protected by the Copyright Act" and preempted under 17 U.S.C. § 301. Dkt. 56 at 4 (dismissing claims).

B. The TAC's New Allegations of Purported Wrongdoing

In the TAC, Plaintiffs assert new, but familiar claims under DMCA and California state law. As to the CMI claim, they now allege that, after downloading the LibGen dataset in April 2023—i.e., after the public release of Llama 1—Meta removed CMI from books included in that dataset "to facilitate training its Llama models by 'cleaning' them for easier 'ingestion' and also to reduce the chance that the models will memorize [CMI]." TAC ¶ 90. According to Plaintiffs, "Meta knew or had reasonable grounds to know that" doing so "would induce, enable, facilitate, or conceal its own copyright infringement or the copyright infringement of others." *Id.* ¶ 108.

Plaintiffs' revamped DMCA claim is not *expressly* tied to the prior theory rejected by the Court: that every Llama output is an infringing derivative of the models' training data. Dkt 56 at 3. But just like the prior claim, the new claim presupposes that "others" have used Llama to generate outputs that infringe Plaintiffs' books, without actually identifying a single example that this ever occurred. That is the same omission that doomed Plaintiffs' earlier DMCA claims. *Id.* As to Plaintiffs' allegation that Meta's removal of CMI aids or conceals Meta's own infringement (TAC ¶ 108), the TAC suggests that this has something to do with memorization and regurgitation of CMI by Meta's LLMs, but the connection between the two is left unexplained.

With respect to Plaintiffs' CDAFA claim, they allege that Meta "knowingly accessed and used data owned by Plaintiffs (*i.e.*, the Infringed Works) without permission in violation of California Penal Code § 502(c)(2)." *Id.* ¶ 111. Plaintiffs' basis for this allegation is Meta's downloading of portions of LibGen and other "shadow library" data, which Plaintiffs allege contain their copyrighted works, via torrent systems and alleged "seeding" of that data, which Plaintiffs claim resulted in Meta "acting as a distribution point for other users of pirated books." *Id.* ¶¶ 43, 46, 87. In other words, Plaintiffs allege that by accessing torrent systems to obtain training data, which Plaintiffs characterize as "pirated, unlicensed, hacked, downloaded, and/or scraped versions

of the Infringed Works," and "sharing, distributing, and/or uploading the same works in the process," Meta violated § 502(c)(2) of CDAFA. *Id.* ¶ 111. These same allegations now also form the basis for Plaintiffs' new distribution theory of direct copyright infringement. *Id.* ¶¶ 87, 101.

III. LEGAL STANDARDS

A. Federal Rule of Civil Procedure 12(b)(1)

The Article III "case or controversy" requirement "ensure[s] that federal courts do not exceed their authority" by limiting their subject matter jurisdiction to cases in which plaintiffs have standing. *Spokeo, Inc. v. Robins*, 578 U.S. 330, 338 (2016). To plead Article III standing, "the plaintiff must 'clearly ... allege facts demonstrating'" that it has suffered an injury in fact, i.e. "'an invasion of a legally protected interest' that is 'concrete and particularized' and 'actual or imminent, not conjectural or hypothetical.'" *Id.* at 338–339 (quoting *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560 (1992)). Where a defendant challenges a plaintiff's Article III standing based on the insufficiency of the allegations in the complaint, a court applies the same standard of review it applies on a motion to dismiss under Rule 12(b)(6). *See Stevens v. Harper*, 213 F.R.D. 358, 370 (E.D. Cal. 2002) ("On a motion to dismiss for lack of standing, ... the court is not obliged to accept allegations of future injury which are overly generalized, conclusory, or speculative" and "[i]n the absence of such specific factual allegations, the court may not assume that jurisdiction exists by 'embellishing otherwise deficient allegations of standing.'" (cleaned up); *see also Colony Cove Props., LLC v. City of Carson*, 640 F.3d 948, 955 (9th Cir. 2011) (applying *Iqbal*'s standards to a motion to dismiss for lack of subject matter jurisdiction and for failure to state a claim).

B. Federal Rule of Civil Procedure 12(b)(6)

"Dismissal under Rule 12(b)(6) is proper when the complaint either (1) lacks a cognizable legal theory or (2) fails to allege sufficient facts to support a cognizable legal theory." *Somers v. Apple, Inc.*, 729 F.3d 953, 959 (9th Cir. 2013). To avoid dismissal, a complaint must plead "enough facts to state a claim to relief that is plausible on its face." *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570 (2007). The court need not "accept as true allegations that contradict … matters properly subject to judicial notice, or allegations that are merely conclusory, unwarranted deductions of fact, or unreasonable inferences." *Daniels-Hall v. Nat'l Educ. Ass'n*, 629 F.3d 992, 998 (9th Cir. 2010);

Dahlia v. Rodriguez, 735 F.3d 1060, 1076 (9th Cir. 2013). "Where a complaint pleads facts that are merely consistent with a defendant's liability, it stops short of the line between possibility and plausibility of entitlement to relief." *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) (cleaned up).

"Because [the] CDAFA claim sounds in fraud, it is subject to Rule 9(b) pleading standards." *Nowak v. Xapo, Inc.*, 2020 WL 6822888, at *5 (N.D. Cal. Nov. 20, 2020) (cleaned up). Plaintiffs must "state with particularity the circumstances regarding fraud," Fed. R. Civ. P. 9(b), which means "stat[ing] precisely the time, place, and nature of the misleading statements, misrepresentations, and specific acts of fraud." *Kaplan v. Rose*, 49 F.3d 1363, 1370 (9th Cir. 1994).

IV. ARGUMENT

A. The TAC Fails to State a Claim Under § 1202(b)(1) of the DMCA (Claim 2)

The TAC alleges that Meta violated Section 1202(b)(1) by removing CMI from the asserted works to conceal Meta's infringement of those works and (unidentified) infringement of others. The Court expressed its skepticism regarding this claim at the hearing on Plaintiffs' Motion for Leave to Amend. Jan. 9, 2025 Tr. at 4 ("I'm maybe a little skeptical about the merits of the [DMCA] claim, but…we could deal with that on a motion to dismiss"). Indeed, this theory doesn't add up. Plaintiffs allege no injury; nor do they plausibly articulate *how* removal of CMI from books prior to training could conceal or facilitate infringement, let alone that Meta acted with the requisite intent. Each of these omissions is fatal to Plaintiffs' DMCA claim.

1. Plaintiffs Lack Standing Due to Failure to Allege a Cognizable Injury

Only persons "injured by a violation of section 1201 or 1202" are authorized to "bring a civil action" under those provisions. 17 U.S.C. § 1203(a). In addition, Plaintiffs "bear the burden of demonstrating that they have [Article III] standing." *TransUnion LLC v. Ramirez*, 594 U.S. 413, 430–31 (2021); *Steele v. Bongiovi*, 784 F. Supp. 2d 94, 97–98 (D. Mass. 2011) (dismissing DMCA claim for failure to plead injury); *Alan Ross Mach. Corp. v. Machinio Corp*, 2019 WL 1317664, at *4 (N.D. Ill. Mar. 22, 2019) (same). As "standing is not dispensed in gross," a plaintiff must establish standing "for each claim he seeks to press and for each form of relief that is sought." *Davis v. FEC*, 554 U.S. 724, 734 (2008) (cleaned up). Standing requires that "the plaintiff's injury in fact be concrete—that is, real and not abstract." *TransUnion*, 594 U.S. at 424, 426 (courts

"cannot treat an injury as concrete for Article III purposes based only on Congress's say-so") (cleaned up). A plaintiff must allege an injury with a "close historical or common-law analogue" such as "physical, monetary, or cognizable intangible harm traditionally recognized as providing a basis for a lawsuit in American courts," to plead a concrete injury. *Id.* at 424, 427.

Here, Plaintiffs fail to plausibly plead any cognizable injury stemming from Meta's alleged removal of CMI from Plaintiffs' asserted works. The TAC contains only a single, threadbare assertion that "Plaintiffs and the Class *have* been injured by Meta's removal of the CMI from the Infringed Works[.]" TAC ¶ 109 (emphasis added). However, "the mere removal of identifying information from a copyrighted work" is ***not***, in and of itself, a cognizable injury. *Raw Story Media, Inc. v. OpenAI, Inc.*, 2024 WL 4711729, at *4 (S.D.N.Y. Nov. 7, 2024) (citing *TransUnion*, 594 U.S. at 424). In other words, Plaintiffs' conclusory assertion that it was injured by CMI removal – devoid of any facts as to how or why Plaintiffs were supposedly injured – is insufficient to establish standing under either Article III or the Copyright Act. *Iqbal*, 556 U.S. at 678.

The Supreme Court's *TransUnion* decision is instructive. There, TransUnion incorrectly labeled the plaintiffs as potential terrorists in its credit files, violating the Fair Credit Reporting Act. 594 U.S. at 418–21. The Court held that although all of the plaintiffs' credit files were affected by TransUnion's statutory violation, only those whose information was actually disseminated had Article III standing. *Id.* at 442. Specifically, those individuals articulated a concrete injury in the form of "reputational harm." *Id.* at 432. In contrast, most plaintiffs' reports were never circulated. For them, the Court reasoned that the "mere presence of an inaccuracy in an internal credit file, if it is not disclosed to a third party, causes no concrete harm" sufficient for standing. *Id.* at 434.

The standing principles articulated in *TransUnion* were recently applied by the district court in *Raw Story*, which addressed removal of CMI in the context of training OpenAI's LLMs, and resulted in dismissal under Rule 12(b)(1). In *Raw Story*, plaintiffs alleged that OpenAI had ingested their "copyrighted works of journalism," "stripped" them of their CMI, and "input" the content into ChatGPT's repository in violation of Section 1202(b)(1). *See* 2024 WL 4711729, at *1. As here, however, the complaint did not allege any specific instances of ChatGPT outputting the plaintiffs' work without CMI; the plaintiffs only suggested "a substantial likelihood" of this possibility. *Id.*

at *1. The district court explained that it was not enough for the plaintiffs to allege that their "copyrighted works (absent CMI) were used to train an AI-software program and remain in [a] repository" (*id.* at *4–5), as this failed to identify "any *actual* adverse effects stemming from this alleged DMCA violation." *Id.* Additionally, after reviewing the legislative history of the DMCA, the court expressed skepticism that "mere removal of identifying information from a copyrighted work—absent dissemination— has *any* historical or common-law analogue," let alone the close analogue required. *Id.* at *4 (emphasis in original). Without such a "close relationship to a harm traditionally recognized as providing a basis for a lawsuit in American courts," *id.* at *2–4, the Section 1202(b)(1) claim could not survive dismissal under *TransUnion*. *Id.*

Plaintiffs' Section 1202(b)(1) claim closely parallels the one dismissed in *Raw Story* and requires the same result. Like OpenAI, Meta is alleged to have removed CMI outside of public view, prior to training certain of its models. And like the plaintiffs in *Raw Story*, Plaintiffs do not allege that they suffered harm with a close historical analogue, such as the dissemination of their works with false or omitted CMI. That is, Plaintiffs still do not identify a single instance in which anyone has ever used Llama to reproduce Plaintiff's books without their CMI—the one theory this Court previously explained might support a claim. Dkt. 52 at 23.[2] *See TransUnion*, 594 U.S. at 414 (explaining that "[t]he mere existence of inaccurate information, absent dissemination, traditionally has not provided the basis for a lawsuit in American courts"). Indeed, as the Supreme Court explained in *TransUnion*, "where allegedly inaccurate or misleading information sits in a company database, the plaintiffs' harm is roughly the same, legally speaking, as if someone wrote a defamatory letter and then stored it in her desk drawer," which is to say, ***no harm at all***. *Id.* at 434.[3] Because Plaintiffs have not asserted an injury caused by Meta's alleged removal of CMI, they lack both Article III standing and statutory standing under Section 1203. *See Doe 1 v. Github, Inc.*, 672 F. Supp. 3d 837, 850–52 (N.D. Cal. 2023) (finding the mere chance of "increased risk"

[2] Indeed, Plaintiffs have amended their complaint twice and with discovery closed, have not asserted any claim relating to the output of Llama or that Llama reproduces their books. Moreover, Plaintiffs have admitted the lack of any evidence of loss of books sales due to Llama or its output. *See, e.g.* Dkt. 329-10 (admitting no evidence of loss of books sales due to Llama).

[3] Any argument that Plaintiffs' works are allegedly kept in Meta's datasets (*e.g.*, TAC ¶ 35) is insufficient to establish standing. *Phillips v. U.S. Customs & Border Prot.*, 74 F.4th 986, 992 (9th Cir. 2023) ("Under [*TransUnion*], the retention of records alone [is not] a concrete injury.").

that plaintiffs' information, input into an AI's model "training data," may eventually be outputted with "missing or incorrect attribution [or] copyright notices" "cannot provide standing for [p]laintiffs' damages claims").[4]

2. Plaintiffs Fail to State a Claim under §1202(b)(1)

Even if Plaintiffs could allege standing and injury under Article III and Section 1203, which they cannot, their DMCA claim must still be dismissed. Section 1202(b)(1) of the DMCA prohibits intentional removal of CMI with knowledge or "reasonable grounds to know, that it will induce, enable, facilitate, or conceal" infringement. A plaintiff "must make an affirmative showing, such as by demonstrating a past 'pattern of conduct' or 'modus operandi,' that the defendant was aware or had reasonable grounds to be aware of the probable future impact of its actions." *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 674 (9th Cir. 2018). This requires factual allegations "plausibly showing that the alleged infringer had this required mental state." *Andersen v. Stability AI Ltd.*, 700 F. Supp. 3d 853, 871 (N.D. Cal. Oct. 30, 2023); *Philpot v. Alternet Media, Inc.*, 2018 WL 6267876, at *5 (N.D. Cal. Nov. 30, 2018) (dismissing DMCA claim where plaintiff "fail[ed] to plead any facts showing that [defendant] had the required mental state"). The TAC's conclusory allegations fall short for several reasons.

First, Plaintiffs' theory of "CMI-removal-as-infringement-concealment" is vague and nonsensical. They suggest CMI removal hides Meta's infringement because "the models cannot regurgitate data they are not trained on," TAC ¶ 90, but there's no plausible explanation of how the presence or absence of CMI in training data is supposed to correlate with revealing Meta's alleged infringement. They do not allege that the CMI removal somehow stymied anyone (including Plaintiffs, themselves) from learning what Llama was trained on or that including CMI in the training data would have made this easier to discern. Tellingly, though fact discovery has closed, the TAC does not cite to any documents or deposition testimony supporting Plaintiffs' theory that Meta removed CMI to aid or conceal infringement. Yet, at the same time Plaintiffs acknowledge that removal of CMI *facilitates* training (TAC ¶ 90), which is inconsistent with their theory of

[4] Plaintiffs have appealed a subsequent order in *Github* dismissing this claim again on different grounds to the Ninth Circuit. That appeal is not directed to the decision cited herein.

concealment.

Second, Plaintiffs baldly allege that Meta knew or had reason to know that removal of CMI "would induce, enable, facilitate, or conceal … the copyright infringement *of others*." (*Id.* ¶ 108 (emphasis added). This claim resembles a similar one that the Court previously dismissed. *Compare* Dkt. 1 ¶ 52 (alleging that removal of CMI from training data would facilitate infringement "by concealing the fact that every output from the LLaMA language models" (*i.e.*, outputs generated by others), "is an infringing derivative work") *with* Dkt. 56 at 3 (Order dismissing DMCA claim). Once again, Plaintiffs "identify no instance in which the removal of CMI" has aided or concealed infringement, because they still allege no instance in which Llama has ever been used to generate an infringing output, much less one that infringes Plaintiffs' at-issue works. *Stevens*, 899 F.3d at 675–76. This theory made no sense in 2023 and makes no sense today.

Third, the claim fails because Plaintiffs do not and cannot plausibly allege that Meta intentionally removed CMI from *Plaintiffs'* books with knowledge or reason to know that doing so would conceal Meta's alleged infringement of those books. Rather, all plausible inferences are wholly inconsistent with that theory. "[T]he mental state requirement in Section 1202(b)" has "a more specific application than the universal possibility of encouraging infringement; *specific allegations as to how identifiable infringements 'will' be affected are necessary*." *Id.* at 674–76 (affirming grant of summary judgment to defendant on § 1202(b) claim) (emphasis added). A "plaintiff must provide evidence from which one can infer that future infringement is likely, albeit not certain, to occur as a result of the removal or alteration of CMI." *Id*. at 675; *see Mills v. Netflix, Inc.*, 2020 WL 548558, at *3 (C.D. Cal. Feb. 3, 2020) (dismissing DMCA claim because scienter was inadequately pleaded); *Victor Elias Photography, LLC v. Ice Portal, Inc*., 43 F.4th 1313, 1325 (11th Cir. 2022) ("[T]he statute's plain language requires some identifiable connection between the defendant's actions and the infringement or the likelihood of infringement.").

The allegations in the TAC do not allow such an inference. Meta *voluntarily published* that it used the Books3 dataset (containing text from Plaintiffs' books) to train Llama 1, a fact which

was the premise of the original Complaint.[5] Dkt. 1 ¶¶ 23-30, 38. It is irrelevant whether Meta later removed CMI from other datasets used to train *Llama 3* (not released until April 2024 (*see* TAC ¶ 84)), because by then Plaintiffs already knew that Meta had used a dataset containing text from their books. Further, any datasets used (and any CMI removal) after Plaintiffs' lawsuit was filed (*see id*. ¶¶ 84–88) was subject to disclosure, belying any connection between CMI removal and concealment of purported infringement of Plaintiffs' works. There simply is no plausible basis to conclude that Meta's removal of CMI from pretraining data aided or concealed infringement of Plaintiffs' at-issue works, let alone that Meta's actions were undertaken for this purpose. The claim must be dismissed.

B. The TAC Fails to State a Claim for Violation of the CDAFA (Claim 3)

Plaintiffs' attempt to characterize Meta's torrenting and alleged seeding of portions of LibGen as a violation of CDAFA is akin to trying to fit a square peg into a round hole. CDAFA is primarily an "anti-hacking statute intended to prohibit the unauthorized use of any computer system for improper or illegitimate purpose." *Sunbelt Rentals, Inc. v. Victor*, 43 F. Supp. 3d 1026, 1032 (N.D. Cal. 2014). Plaintiffs claim that by downloading data from the LibGen dataset via bit torrent and allegedly "seeding" that data, Meta violated subsection (c)(2) of the statute, which prohibits "[k]nowingly access[ing] and without permission tak[ing], cop[ying], or mak[ing] use of any data from a computer, computer system, or computer network[.]" Cal. Penal Code § 502(c)(2). This recasting of Plaintiffs' core copyright claim is both preempted and otherwise legally untenable.

1. The Copyright Act Preempts Plaintiffs' CDAFA Claim

The Court has already signaled skepticism of the CDAFA claim, and rightfully so. *See* Jan. 9, 2025 Tr. at 4:9–11 ("it seems like there's probably a pretty good argument that the CDAFA claim

[5] As Plaintiffs know—and have alleged—"in preparing [] dataset[s] for use in training its models, [Meta] removed repetitive data," which included by way of example, the word copyright, the © symbol, excessive new line characters (\n), and "all rights reserved," as well as lines that are at least 100 words long that contain fewer than 8% unique words, among other repetitive content. See, e.g., Dkt. 395-8 at 89798-89799 (listing removal of this repetitive information as one of several "improvements") (emphasis added). According to the TAC, Meta did this by "script" across all books "to facilitate Meta's use of these works as training data for Llama 3" and to avoid "regurgitation" of CMI (*id*. ¶¶ 88–90). These allegations do not even plausibly establish that Meta targeted Plaintiffs' books for CMI removal, much less that it did so with intent to "hide" any purported infringement of those books.

is preempted…"). Under the Copyright Act, "all legal or equitable rights that are equivalent to any of the exclusive rights within the general scope of copyright as specified by section 106 … are governed exclusively by this title," and no person "is entitled to any such right or equivalent right in any such work under the common law or statutes of any State." 17 U.S.C. § 301(a). Courts use a two-part test to assess whether a state law claim is preempted. The court first asks "whether the subject matter of the state law claim falls within the subject matter of copyright as described in 17 U.S.C. §§ 102 and 103." *Maloney v. T3Media, Inc.*, 853 F.3d 1004, 1010 (9th Cir. 2017) (cleaned up). If so, it then assesses "whether the rights asserted under state law are equivalent to the rights contained in 17 U.S.C. § 106." *Laws v. Sony Music Entm't, Inc.*, 448 F.3d 1134, 1138 (9th Cir. 2006). Under this test, the CDAFA claim is preempted.

To begin, the "subject matter" of the CDAFA claim is Plaintiffs' books, which fall squarely within the subject matter of copyright. *See Maloney*, 853 F.3d at 1011. Plaintiffs concede that the "data" at issue in their CDAFA claim consists of the allegedly "Infringed Works," TAC ¶ 111, which are defined as "the *copyrighted books that Meta copied* and used without permission to train Llama, regardless of where or how Meta downloaded or otherwise accessed the books," *id.* ¶ 46 (emphasis added). The Infringed Works also form the basis of Plaintiffs' copyright infringement claim. *See*, *e.g.*, *id.* ¶ 100 ("Meta made copies of the Infringed Works during the training process to develop Llama without Plaintiffs' permission"). Plaintiffs' CDAFA claim thus involves works within the subject matter of copyright—easily satisfying the first prong of the preemption test. *See* 17 U.S.C. § 102 (literary works are the subject matter of copyright); *see also Tremblay v. OpenAI, Inc.*, 2024 WL 3640501, at *2 (N.D. Cal. July 30, 2024) (holding that UCL claim fell under subject matter of copyright where claim involved unauthorized use of plaintiffs' books).

As to the second prong of the test, the rights Plaintiffs assert under the CDAFA claim are "equivalent to the rights contained" in Section 106. *Laws*, 448 F.3d at 1138. "To survive preemption, the state cause of action must protect rights which are qualitatively different from the copyright rights. The state claim must have an extra element which changes the nature of the action." *Maloney* at 1019 (cleaned up). Even if a state law claim has different elements, it may still be preempted if its claims "are part and parcel of the copyright claim." *Laws*, 448 F.3d at 1144.

Plaintiffs allege that Meta violated Section 502(c)(2) by "obtaining pirated, unlicensed, hacked, downloaded, and/or scraped versions of the Infringed Works via bit torrent protocols for use to train Llama models and by sharing, distributing, and/or uploading the same works in the process." TAC ¶ 111; *see id.* ¶¶ 43, 87, 110–114. Essentially, they claim that Meta downloaded (copied) their "Infringed Works" from unauthorized websites and distributed them without their permission. Both acts fall squarely within the exclusive rights afforded by Sections 106(1) and (3) of the Copyright Act. Indeed, this is nothing more than a transparent reframing of their core copyright infringement claim, as illustrated in the table below:

Plaintiffs' CDAFA Claim (TAC ¶ 111):	**Plaintiffs' Copyright Claim (TAC ¶¶ 98–101):**
"By obtaining pirated, unlicensed, hacked, downloaded, and/or scraped versions of the Infringed Works via bit torrent protocols for use to train Llama models and by sharing, distributing, and/or uploading the same works in the process, Defendant knowingly accessed and used data owned by Plaintiffs' (i.e., the Infringed Works) without permission…"	"To train Llama…Meta also downloaded and copied millions of additional pirated books, including the Infringed Works from various shadow libraries…Plaintiffs never authorized Meta to make copies of their Infringed Works…Meta made copies of the Infringed Works during the training process to develop Llama without Plaintiffs' permission. During the download process of LibGen and other shadow libraries…Meta also operated as a distributor of the pirated works…"

Courts, including this Court, routinely find that state-law claims grounded in allegations of copyright infringement are preempted. *See, e.g.*, Dkt. 56 (dismissing Plaintiffs' unfair competition, unjust enrichment, and negligence claims as relying on the same rights contained in the Copyright Act and thus preempted); *Regal Art & Gifts, Inc. v. Fusion Prod., Ltd.*, 2016 WL 454116, at *7 (N.D. Cal. Feb. 5, 2016) (dismissing claims for intentional interference, negligent interference, and unfair competition as preempted where based "wholly on Defendants' allegedly unauthorized copying and distribution" of the plaintiffs' copyrighted work); *see also VBConversions, LLC v. Blueswitch, LLC*, No. 2:25-cv-09372 (C.D. Cal. Mar. 18, 2016), ECF No. 58 (Anderson, J.) (granting motion to dismiss CDAFA claim as preempted by the DMCA).

In their Motion for Leave to Amend, Plaintiffs argued that their CDAFA claim is not preempted because "the focus on unauthorized access to [their…] texts gives the CDAFA statute an 'extra element' that makes it 'qualitatively different' from copyright law." Dkt. 376 at 10 n.6

(citations omitted). However, as Meta then noted, the cases cited by Plaintiffs were readily distinguishable. *See* Dkt. 378 at 13. For example, *Altera Corp. v. Clear Logic, Inc.* dealt with a contractual right of access to "use of the bitstream," ***not*** "reproduction of the software." 424 F.3d 1079, 1089 (9th Cir. 2005). *Capitol Audio Access, Inc. v. Umemoto* involved unauthorized access to payment- and password-protected portions of the plaintiff's website "in direct contravention of the User License and User Agreement." 980 F. Supp. 2d 1154, 1156 (E.D. Cal. 2013). *Grosso v. Miramax Film Corp.* also concerned a separate contractual right. 383 F.3d 965, 968 (9th Cir. 2004). In this case, Meta's copying and alleged distribution (by "seeding") of Plaintiffs' "Infringed Works" (books) without authorization forms the sole factual basis of the CDAFA claim. That is nothing more than a copyright claim and, hence, preempted.[6]

2. The TAC Fails to Plausibly Allege a Violation of Section 502(c)(2)

CDAFA imposes liability if one "[k]nowingly accesses and without permission takes, copies, or makes use of any data from a computer, computer system, or computer network[.]" Cal. Penal Code § 502(c)(2); *United States v. Christensen*, 828 F.3d 763, 789 (9th Cir. 2015). Plaintiffs fail to plausibly allege conduct by Meta that violates § 502(c)(2), let alone meets the heightened pleading standard under Rule 9(b).

First, there is no legal basis for Plaintiffs' attenuated theory of liability under CDAFA. The statute requires that the claimant be "the owner or lessee of the computer, computer system, computer network, computer program, or data" allegedly accessed without permission. Cal. Pen. Code § 502(e)(1).[7] But Plaintiffs do not allege this in the TAC, because they cannot. Instead, at most, Plaintiffs allege that Meta downloaded datasets created by third parties from computer systems operated by third parties. This is too far removed to support a CDAFA claim. *See Claridge v. RockYou, Inc.*, 785 F. Supp. 2d 855, 863 (N.D. Cal. 2011) ("[I]t is less than clear that [CDAFA] is meant to subject individuals or entities to liability who took no active role in tampering with, or

[6] Copyright infringement itself requires proof of "access" and "substantial similarity" in protectable expression. *Transgo, Inc. v. Ajac Transmission Parts Corp.*, 768 F.2d 1001, 1018 (9th Cir. 1985).

[7] Section 502(a) explains that the statute's intent was to "expand the degree of protection afforded to individuals, businesses, and governmental agencies from tampering, interference, damage, and unauthorized access to ***lawfully created computer data and computer systems***." Cal. Pen. Code § 502(a) (emphasis added).

in gaining unauthorized access to computer systems."). Indeed, Meta is not aware of a single case—and Plaintiffs have yet to cite one—in which a CDAFA claimant did not own or control the computer systems or data repositories allegedly accessed.

As they did earlier (Dkt. 376 at 10), Plaintiffs may point to *Biden v. Ziegler*, where the court rejected an argument that CDAFA requires a defendant to access a physical computer or device belonging to or controlled by the plaintiff. 737 F. Supp. 3d 958, 976 (C.D. Cal. 2024). But this case is readily distinguishable on its facts. In *Biden,* there was no question that the plaintiff owned and controlled the set of files at issue, which consisted of his emails, photos, videos, and records from a copy of his laptop hard drive or from his encrypted phone backup stored by him on the cloud. *Id.* at 966. Further, the plaintiff there alleged – unlike here – that the defendants "*used his passwords to access password-protected files* and ignored prelitigation demands to cease and desist[]." *Id*. at 976 (emphasis added). Similarly, in *West v. Ronquillo-Morgan* (previously cited at Dkt. 377 at 10), the court denied a motion to dismiss a CDAFA claim, holding that "the data was obtained from Plaintiff's computer and therefore the data was 'hers.'" 526 F. Supp. 3d 737, 746 (C.D. Cal. 2020).

Here, the TAC contains no allegations regarding how Plaintiffs' copyrighted works made it into the datasets that Meta downloaded, much less that Plaintiffs created those datasets or owned or controlled the websites or computers on which those datasets were maintained. Moreover, the computer systems and datasets that Meta accessed were publicly available. This is a far cry from cases like *Christensen*, in which the defendants paid SBC employees to access and share the victims' personal phone conversations on SBC's systems. *Christensen*, 828 F.3d at 789. Plaintiffs have thus failed to allege any computer fraud perpetrated upon them, either generally or with the particularity required under Rule 9(b)'s heightened pleading standard. *Kaplan*, 49 F.3d at 1370. Indeed, the TAC specifically acknowledges that the datasets Meta used were compiled by third party "websites," not Plaintiffs or Meta, *see, e.g.*, TAC ¶ 43, taking this claim outside of CDAFA's plausible ambit. Were it otherwise, a copyright defendant would be subject to potential liability under a state criminal statute whenever it downloads copyrighted content from the internet without

permission—something no court has ever sanctioned. This Court should decline to read into CDAFA the unprecedented breadth required to sustain Plaintiffs' application of it here.

Second, Plaintiffs do not allege the requisite damage or loss to support their CDAFA claim. Under the statute, compensatory damages include "any expenditure reasonably and necessarily incurred by the owner or lessee to verify that a computer system, computer network, computer program, or data was or was not altered, damaged, or deleted by the access." Cal. Pen. Code § 502(e)(1). As such, CDAFA "contemplates some damage to the computer system, network, program, or data[.]" *Heiting v. Taro Pharms. USA, Inc.*, 709 F. Supp. 3d 1007, 1021 (C.D. Cal. 2023); *see also Pratt v. Higgins*, 2023 WL 4564551, at *9 (N.D. Cal. July 17, 2023) ("In the context of a § 502 violation, 'loss' has been defined to encompass costs related to fixing a computer, lost revenue, or other consequential damages incurred due to an interruption of computer services."). The TAC, however, contains no such allegations of damage.

In conclusory fashion, Plaintiffs allege that they "have been harmed in an amount to be determined at trial, including but not limited to lost royalties, reputational damages, and other consequential losses," which is insufficient. TAC ¶ 114. *See Pratt,* 2023 WL 4564551, at *9 (holding plaintiff's request for "damages in a sum to be determined at trial" insufficient).[8] Even "loss of the right to control…, the loss of the value of…, and the loss of the right to protection of" their data, are not types of loss covered by the CDAFA. *See Cottle v. Plaid Inc.*, 536 F. Supp. 3d 461, 488 (N.D. Cal. 2021); *Doe v. Meta Platforms, Inc.*, 690 F. Supp. 3d 1064, 1081-83 (N.D. Cal. 2023) (finding allegations that protected information was diminished in value did not qualify as damage or loss under CDAFA). Plaintiffs' CDAFA claim separately fails on this basis.

V. CONCLUSION

For the foregoing reasons, Meta asks the Court to dismiss with prejudice the Section 1202(b)(1) and CDAFA claims.

[8] To the extent Plaintiffs are asserting lost "royalties," that only further underscores that this claim should be preempted under copyright.

Dated: January 31, 2025

COOLEY LLP

By: _____

Bobby Ghajar
Kathleen Hartnett
Mark Weinstein
Phillip Morton
Judd Lauter
Liz Stameshkin
Matthew Brigham
Colette Ghazarian

CLEARY GOTTLIEB STEEN & HAMILTON LLP

Angela L. Dunning

Attorneys for Defendant
META PLATFORMS, INC.

Full Counsel List

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

COOLEY LLP
MATTHEW BRIGHAM (191428)
(mbrigham@cooley.com)
JUAN PABLO GONZALEZ (334470)
(jgonzalez@cooley.com)
3175 Hanover Street
Palo Alto, CA 94304-1130
Telephone: (650) 843-5000