

No. 24-7700

---

---

**United States Court of Appeals  
for the  
Ninth Circuit**

---

J. DOE 1, et al.,  
*Plaintiffs-Appellants,*

*v.*

GITHUB, INC., et al.,  
*Defendants-Respondents.*

---

On Appeal from the United States District Court  
for the Northern District of California  
No. 4:22-cv-06823, Hon. Jon S. Tigar

---

**APPELLANTS' REPLY BRIEF**

---

JOSEPH SAVERI  
LAW FIRM, LLP  
Joseph R. Saveri  
Christopher K.L. Young  
601 California St., 1505  
San Francisco, CA 94108  
(415) 500-6800  
jsaveri@saverilawfirm.com  
cyoung@saverilawfirm.com

Matthew Butterick  
1920 Hillhurst Ave.  
Ste. 406  
Los Angeles, CA 90027  
(323) 968-2632  
mb@buttericklaw.com

BOIES SCHILLER FLEXNER LLP  
Jesse Panuccio  
Evan Ezray  
Daniel Morales  
401 East Las Olas Blvd.  
Fort Lauderdale, FL 33301  
jpanuccio@bsfllp.com  
eezray@bsfllp.com

Maxwell V. Pritt  
44 Montgomery St., 41st Fl.  
San Francisco, CA 94104  
(415) 293-6800  
mpritt@bsfllp.com

*Counsel for Plaintiffs-Appellants and Proposed Class*

## TABLE OF CONTENTS

TABLE OF AUTHORITIES.....	iv
INTRODUCTION.....	1
ARGUMENT .....	3
I.    DEFENDANTS ABANDON THE IDENTICALITY REQUIREMENT THEY ADVANCED BELOW .....	3
A.    Defendants Concede the Only Issue on Appeal .....	3
B.    Defendants Waived Their New “Weighty Factor” Theory.....	4
II.   DEFENDANTS’ NEW THEORY FAILS ON THE MERITS.....	6
A.    Plaintiffs Satisfy Defendants’ Requirement of Alleging Removal from an Existing Work.....	7
B.    Defendants’ Myopic Focus on Identicality as an Evidentiary Factor Contradicts Their Concessions and Fails on the Merits .....	14
C.    Plaintiffs Do Not Ask for an Attribution Right Under the DMCA .....	20
III.  PLAINTIFFS HAVE STANDING.....	23
A.    Plaintiffs’ Injuries Are Concrete and Particularized.....	23
1.    CMI Removal Is an Existing, Concrete Injury.....	24
2.    These Injuries Are Redressable, Not Moot.....	27

3.	Plaintiffs Have Shown Substantial Risk of Future Harm by Concrete Evidence, Not Speculation.....	29
B.	Plaintiffs’ Injuries Are Not Manufactured .....	37
IV.	PLAINTIFFS ADEQUATELY ALLEGE SCIENTER .....	40
	CONCLUSION .....	45

## TABLE OF AUTHORITIES

Page(s)

### OTHER AUTHORITIES

<i>Aardwolf Indus., LLC v. Abaco Machines USA, Inc.</i> , 2017 WL 10350547 (C.D. Cal. Nov. 13, 2017) .....	10
<i>ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.</i> , 667 F.Supp.3d 411 (S.D. Tex. 2023).....	21
<i>APL Microscopic, LLC v. Steenblock</i> , 2022 WL 4830687 (9th Cir. Oct. 3, 2022) .....	42
<i>Ariz. Students’ Ass’n v. Ariz. Bd. of Regents</i> , 824 F.3d 858 (9th Cir. 2016) .....	12
<i>Bain v. Film Indep., Inc.</i> , 2022 WL 17592422 (9th Cir. Dec. 13, 2022) .....	41
<i>Bell Atl. Corp. v. Twombly</i> , 550 U.S. 544 (2007) .....	13
<i>Brown v. Stroud</i> , 2011 WL 2600661 (N.D. Cal. June 30, 2011).....	43
<i>Clapper v. Amnesty Int’l USA</i> , 568 U.S. 398 (2013) .....	38, 39
<i>Davidson v. Kimberly-Clark Corp.</i> , 889 F.3d 956 (9th Cir. 2018) .....	28
<i>DeFries v. Union Pac. R.R. Co.</i> , 104 F.4th 1091 (9th Cir. 2024).....	6, 41
<i>Doe v. Hochul</i> , 139 F.4th 165 (2d Cir. 2025) .....	32
<i>E.T. v. Paxton</i> , 41 F.4th 709 (5th Cir. 2022).....	33

<i>F.T.C. v. Network Servs. Depot, Inc.</i> , 617 F.3d 1127 (9th Cir. 2010) .....	42
<i>Falkner v. General Motors</i> , 393 F.Supp.3d 927 (C.D. Cal. 2018).....	16
<i>FDA v. All. for Hippocratic Med.</i> , 602 U.S. 367 (2024) .....	32
<i>FEC v. Cruz</i> , 596 U.S. 289 (2022) .....	37, 39, 40
<i>Fischer v. Forrest</i> , 968 F.3d 216 (2d Cir. 2020).....	25, 28
<i>Food &amp; Water Watch, Inc. v. Vilsack</i> , 808 F.3d 905 (D.C. Cir. 2015).....	30
<i>Friedman v. Live Nation Merch., Inc.</i> , 833 F.3d 1180 (9th Cir. 2016) .....	10, 16, 18, 24, 27, 42
<i>Galaria v. Nationwide Mut. Ins. Co.</i> , 663 F. App'x 384 (6th Cir. 2016) .....	31, 35
<i>GC2 Inc. v. Int'l Game Tech.</i> , 391 F.Supp.3d 828 (N.D. Ill. 2019) .....	15, 16, 18
<i>Harrington v. Pinterest, Inc.</i> , 2021 WL 4033031 (N.D. Cal. Sept. 3, 2021) .....	42, 44
<i>Havens Realty Corp. v. Coleman</i> , 455 U.S. 363 (1982) .....	37
<i>In re Ward</i> , 21 F.3d 1119 (9th Cir. 1994) .....	14
<i>In re Zappos.com, Inc.</i> , 888 F.3d 1020 (9th Cir. 2018) .....	31
<i>Izmo, Inc. v. Roadster, Inc.</i> , 2019 WL 13210561 (N.D. Cal. Mar. 26, 2019).....	10, 11, 43

<i>Jacobsen v. Katzer</i> , 535 F.3d 1373 (Fed. Cir. 2008).....	26
<i>Johnson v. City of Shelby, Miss.</i> , 574 U.S. 10 (2014) .....	7
<i>Kadrey v. Meta Platforms, Inc.</i> , 2025 WL 744032 (N.D. Cal. Mar. 7, 2025).....	9, 25, 44
<i>Knox v. Serv. Emps. Int’l Union</i> , 567 U.S. 298 (2012) .....	29
<i>Lehrman v. Lovo, Inc.</i> , 2025 WL 1902547 (S.D.N.Y. July 10, 2025) .....	9, 11
<i>Lewert v. P.F. Chang’s China Bistro, Inc.</i> , 819 F.3d 963 (7th Cir. 2016) .....	31
<i>Loeb-Defever v. Mako, LLC</i> , 2023 WL 5611042 (5th Cir. Aug. 30, 2023) .....	42
<i>Los Angeles v. Lyons</i> , 461 U.S. 95 (1983) .....	35
<i>Lujan v. Defs. of Wildlife</i> , 504 U.S. 555 (1992) .....	23, 32
<i>M.D. Russell Constr., Inc. v. Consol. Staffing, Inc.</i> , 2023 WL 8798086 (4th Cir. Dec. 20, 2023) .....	14
<i>Magadia v. Wal-Mart Assocs., Inc.</i> , 999 F.3d 668 (9th Cir. 2021) .....	28, 29
<i>Malik v. DHS</i> , 78 F.4th 191 (5th Cir. Aug. 15, 2023) .....	29
<i>Mango v. BuzzFeed, Inc.</i> , 970 F.3d 167 (2d Cir. 2020).....	41
<i>Massachusetts v. HHS</i> , 923 F.3d 209 (1st Cir. 2019).....	32

<i>Mayfield v. United States</i> , 599 F.3d 964 (9th Cir. 2010) .....	28, 34
<i>Microsoft Corp. v. AT&amp;T Corp.</i> , 550 U.S. 437 (2007) .....	19
<i>N.Y. Times Co. v. Microsoft Corp.</i> , 777 F.Supp.3d 283 (S.D.N.Y. 2025) .....	9
<i>Nelsen v. King County</i> , 895 F.2d 1248 (9th Cir. 1990) .....	33, 34
<i>New Parent World, LLC v. True To Life Prods. Inc.</i> , 2024 WL 4277865 (D. Ariz. Sept. 24, 2024).....	21
<i>Oracle Int’l Corp. v. Rimini St., Inc.</i> , 2023 WL 4706127 (D. Nev. July 24, 2023) .....	10, 18
<i>Overbey v. Mayor of Baltimore</i> , 930 F.3d 215 (4th Cir. 2019) .....	27, 29
<i>Philpot v. Alternet Media, Inc.</i> , 2018 WL 6267876 (N.D. Cal. Nov. 30, 2018) .....	42
<i>Pinkert v. Schwab Charitable Fund</i> , 49 F.4th 1051 (9th Cir. 2022).....	35
<i>Quintara Biosciences, Inc. v. Ruifeng Biztech, Inc.</i> , 2025 WL 2315671 (9th Cir. Aug. 12, 2025) .....	9
<i>Raw Story Media, Inc. v. OpenAI, Inc.</i> , 756 F.Supp.3d 1 (S.D.N.Y. 2024) .....	34
<i>Real World Media, LLC v. Daily Caller, Inc.</i> , 744 F.Supp.3d 24 (D.D.C. 2024) .....	21
<i>Robins v. Spokeo, Inc.</i> , 867 F.3d 1108 (9th Cir. 2017) .....	28
<i>Satanic Temple v. Labrador</i> , 2025 WL 2302188 (9th Cir. Aug. 11, 2025) .....	32

<i>Skinner v. Switzer</i> , 562 U.S. 521 (2011) .....	12
<i>Sonterra Cap. Master Fund Ltd. v. UBS AG</i> , 954 F.3d 529 (2d Cir. 2020).....	13
<i>Spokeo, Inc. v. Robins</i> , 578 U.S. 330 (2016) .....	33
<i>Stevens v. CoreLogic</i> , 899 F.3d 666 (9th Cir. 2018) .....	40, 41, 43, 44
<i>Summers v. Earth Island Inst.</i> , 555 U.S. 488 (2009) .....	31
<i>Susan B. Anthony List v. Driehaus</i> , 573 U.S. 149 (2014) .....	30
<i>Swierkiewicz v. Sorema</i> , 534 U.S. 506 (2002) .....	7, 15, 41
<i>Tennessee Conf. of NAACP v. Lee</i> , 139 F.4th 557 (6th Cir. 2025).....	33
<i>Terracom v. Valley Nat. Bank</i> , 49 F.3d 555 (9th Cir. 1995) .....	5
<i>The Intercept Media, Inc. v. OpenAI, Inc.</i> , 767 F.Supp.3d 18 (S.D.N.Y. 2025) .....	9, 25, 27, 36, 40
<i>TransUnion LLC v. Ramirez</i> , 594 U.S. 413 (2021) .....	25
<i>Tremblay v. OpenAI, Inc.</i> , 716 F.Supp.3d 772 (N.D. Cal. 2024) .....	42
<i>United States v. Hernandez</i> , 37 F.4th 1316 (7th Cir. 2022).....	14
<i>Universal City Studios, Inc. v. Corley</i> , 273 F.3d 429 (2d Cir. 2001).....	19, 23

<i>Victor Elias Photography, LLC v. Ice Portal, Inc.</i> , 43 F.4th 1313 (11th Cir. 2022).....	42
<i>Warren v. Fox Family Worldwide, Inc.</i> , 328 F.3d 1136 (9th Cir. 2003) .....	23
<i>Zuma Press, Inc. v. Getty Images, Inc.</i> , 845 F. App'x 54 (2d Cir. 2021).....	41

**STATUTES**

17 U.S.C. § 101 .....	22
17 U.S.C. § 106 .....	26
17 U.S.C. § 1202 .....	<i>passim</i>
17 U.S.C. § 1203 .....	29

**RULES**

Fed. R. Civ. P. 8(a).....	9, 12
Fed. R. Civ. P. 9(b).....	42

**LEGISLATIVE MATERIALS**

S. Rep. No. 105-190 (1998) .....	25
----------------------------------	----

## INTRODUCTION

Microsoft, OpenAI, and GitHub promote their commercial product as a “transformative” innovation that harnesses the benefits of open source’s collaborative spirit. But in their pursuit of market dominance and profit, Defendants trample the very property rights that make open source possible. This species of argument is not new: Purveyors of new technology often claim that application of traditional legal principles to their wares will stifle innovation and harm society—from railroads claiming tort immunity in the 1800s to file-sharing networks eschewing copyright in the early 2000s. These claims proved false then and remain false today. AI is perfectly able to thrive while respecting copyright law. Indeed, Defendants recently signed licensing agreements with major companies like News Corp., AP, and Reddit.<sup>1</sup>

Suffice it to say that Plaintiffs will demonstrate in the district court that there is no valid defense for the violation of their copyrights here. But, despite Defendants’ efforts to litigate the entire case in this interlocutory appeal, the narrow question presented is straightforward

---

<sup>1</sup> <https://variety.com/2024/digital/news/news-corp-openai-licensing-deal-1236013734/>; <https://www.telegraph.co.uk/business/2024/01/07/openai-warns-copyright-crackdown-could-doom-chatgpt/>.

because Defendants concede the district court erred. The district court adopted a strict identity requirement, and Defendants now admit identity is not required. This Court need only accept this concession and reverse.

Having abandoned a strict identity requirement, Defendants advance a new theory for the first time on appeal, recasting identity as a “weighty factor” for determining whether CMI was removed. But this new theory is not properly before this Court; and even if it were, it fails on the merits.

Defendants, therefore, grasp for alternative grounds to affirm. They argue Plaintiffs lack standing despite the concrete injury from the completed removal of CMI, and despite admitting their models frequently regurgitate licensed code (including Plaintiffs’) without CMI. Defendants next baselessly accuse Plaintiffs of manufacturing injuries by demonstrating how Copilot reproduces their code—even though the Supreme Court permits such testing to expose ongoing violations. Defendants then argue that even if they *did* remove CMI, Plaintiffs *still* cannot seek an injunction to prevent present harm from Defendants’ ongoing reproductions. Finally, Defendants demand a heightened

pleading standard for scienter that would require proving actual infringement, which would eviscerate the statute’s prophylactic function. None of these arguments salvage the district court’s erroneous dismissal.

## ARGUMENT

### I. DEFENDANTS ABANDON THE IDENTICALITY REQUIREMENT THEY ADVANCED BELOW

#### A. Defendants Concede the Only Issue on Appeal

At Defendants’ urging,<sup>2</sup> the district court dismissed Plaintiffs’ DMCA claims based on a strict “identity requirement,” holding that even “near-identical cop[ies]” or “functional equivalent[s]” are “not sufficient” for liability. 3-ER-259-60 (“Even where the underlying works are similar, ... no DMCA violation exists unless the works are identical.”); 1-ER-6-7. The district court certified, and this Court granted, interlocutory appeal to resolve only that question: “Is liability under § 1202(b) ... restricted *solely* to the removal or alteration of [CMI] from an *identical* copy of a work?” 2-ER-26 (emphasis added); 2-ER-47.

Defendants now join Plaintiffs in answering, “No.” The district court’s categorical rule—that only identical copies support DMCA

---

<sup>2</sup> 1-ER-7 (“Defendants argue that ... ‘the SAC does not identify even a single example of Copilot producing an identical copy of any work.’ The Court agrees.”); 3-ER-174; 3-ER-317; 1-FER-52; 1-FER-13.

liability—was wrong. Defendants concede that a “copy of the original” need not be “completely identical.” OAB.45;<sup>3</sup> GAB.46.<sup>4</sup> They agree that “Plaintiffs’ arguments rule out a theory that a defendant only violates section 1202(b) by removing or altering CMI on identical ‘copies,’” because the statute “does not require identity,” OAB.47, and thus “lack of identity” cannot “be treated as a rule unto itself,” GAB.51. They further concede that superficial modifications do not automatically preclude DMCA liability. GAB.50-51; OAB.46-47. These concessions should resolve this appeal.

### **B. Defendants Waived Their New “Weighty Factor” Theory**

Having secured dismissal based on an indefensible strict identity requirement, Defendants pivot to a new theory on appeal. They recharacterize “identity” as a “weighty factor” or strong

---

<sup>3</sup> OpenAI’s brief is cited as “OAB,” and GitHub/Microsoft’s as “GAB.”

<sup>4</sup> OAB.41 (“identity” is “not dispositive”); GAB.3 (there is no “freestanding, unbending requirement that a defendant’s ... end-product must be a dead-literal match for the plaintiff’s original”); GAB.46 (rejecting “rigid identity requirement”).

indication that helps determine whether CMI was “removed.” GAB.51; OAB.41-42.

But Defendants cannot raise this new theory for the first time on appeal. *Terracom v. Valley Nat. Bank*, 49 F.3d 555, 559 (9th Cir. 1995) (“this court will not consider contentions which were not originally brought before the district court”). In their motions to dismiss, Defendants advanced a single position: “§ 1202(b) claims lie *only* when CMI is removed or altered from an *identical copy*.” 3-ER-317 (emphasis added). They argued categorically that “Section 1202(b) claims require that copies be identical.” 1-FER-52. They did not suggest that identicality was merely “a weighty factor,” GAB.51, or that it “helps courts identify” removal, OAB.41-42. Instead, they maintained “CMI-removal claims” could *never* be “based on mere excerpts of works” because “*only* removal of CMI from a *complete copy* suffices.” 1-FER-13 (emphasis added); 3-ER-171-77; 1-FER-51-55; 1-FER-26-31; 2-FER-77-86. The district court adopted Defendants’ categorical position, dismissing Plaintiffs’ DMCA claims because the Complaint “failed to allege the removal of CMI from an identical copy.” 3-ER-174. The district

court did not consider—because it was not before it—Defendants’ new legal theory.<sup>5</sup>

“This court is one of review, not first view.” *DeFries v. Union Pac. R.R. Co.*, 104 F.4th 1091, 1109 (9th Cir. 2024) (quotations omitted). Defendants’ new theory, that non-identity is a strong indication CMI was not removed, raises factual questions about how Copilot operates. GAB.50-51; OAB.50. Such fact-intensive inquiries are inappropriate for appellate resolution in the first instance. *DeFries*, 104 F.4th at 1109 (declining to address “not [] purely legal” issue).

## II. DEFENDANTS’ NEW THEORY FAILS ON THE MERITS

Even if Defendants had not waived their new arguments, they would still fail on the merits. First, Plaintiffs and Defendants agree that section 1202(b) is about removing CMI from another’s work, and Plaintiffs’ allegations show Defendants did just that. Defendants claim Plaintiffs do not “advance a standard of their own.” GAB.52. But Plaintiffs’ standard is simply what the statute requires: plausible allegations that Defendants removed CMI. There is no special pleading

---

<sup>5</sup> Defendants assert the district court did not really apply a strict identity argument below. GAB.46. That is not how the district court understood its own order, considering how it formulated the certified question. 2-ER-47.

standard for DMCA claims. *See Johnson v. City of Shelby, Miss.*, 574 U.S. 10, 11 (2014) (“heightened pleading standard” would conflict with the Federal Rules). Second, Defendants nonetheless try to fashion such a pleading standard by repackaging identity as a “dispositive factor.” GAB.23. But this novel, atextual proxy for CMI removal “is an evidentiary standard, not a pleading requirement.” *Swierkiewicz v. Sorema N.A.*, 534 U.S. 506, 510 (2002).

**A. Plaintiffs Satisfy Defendants’ Requirement of Alleging Removal from an Existing Work**

1. Plaintiffs and Defendants agree that section 1202(b) prohibits removing or altering CMI from “existing works.” OAB.32-37; GAB.52-54, 35; OB.28. But Defendants mischaracterize the SAC as alleging merely that Defendants’ “AI tools outputted similar code to Plaintiffs’ without attribution.” OAB.51; GAB.54 (“no allegations suggest that Copilot’s outputs *remove* CMI from a copy of the original”). This distorts Plaintiffs’ allegations: that Defendants systematically copied billions of lines of code and stripped the code’s CMI to conceal when their commercial AI product redistributes that code to millions of users. OB.9-14, 48-53. This is precisely the “affirmative act” of removal that Defendants now acknowledge violates section 1202(b), GAB.51—what they call a “mine-

run case.” OAB.45. Defendants did far “more than merely omit[] or leav[e] out CMI.” OAB.35. They copied existing files containing Plaintiffs’ code with CMI and intentionally deleted that CMI, either at the front-end when processing that data for training, or at the back-end when an output was generated.

Defendants argue that “Codex and Copilot did not ‘remove’ or ‘alter’ CMI.” OAB.4. Instead, they say Copilot “*creates new content*,” OAB.11, that merely “resemble[s]” Plaintiffs’ code, GAB.12. But that mischaracterizes how their AI models work. Defendants admit they trained their models on code taken from GitHub’s public repositories—which included CMI. 3-ER-205, 208 ¶¶ 94-95, 109. The models then reproduce this code in response to user prompts—without CMI. 3-ER-208-10, 217-21 ¶ 108-16, 133-40. When Copilot emits excerpts that match someone’s code, it is not creating something new; it is removing the CMI before regurgitating those excerpts, 3-ER-207-08, 222 ¶¶ 104-05, 141-49. GitHub’s duplicate-detection feature proves this. The feature identifies these violations and even provides links to the missing CMI—a clear admission the models are reproducing someone else’s code rather than creating any “new work.” 3-ER-221 ¶ 139.

Copilot even emitted excerpts of Plaintiffs’ code without its CMI that had Doe 1’s actual “comments in the code”—i.e., Doe’s own expressive commentary. 3-ER-121. This proves Defendants copied Plaintiffs’ entire code files (code, CMI, and extraneous comments) and removed CMI before reproducing the rest. Courts have permitted similar section 1202(b) claims to proceed on much less evidence. *E.g.*, *Kadrey v. Meta Platforms, Inc.*, 2025 WL 744032, at \*2 (N.D. Cal. Mar. 7, 2025); *The Intercept Media, Inc. v. OpenAI, Inc.*, 767 F.Supp.3d 18, 30-31 (S.D.N.Y. 2025); *N.Y. Times Co. v. Microsoft Corp.*, 777 F.Supp.3d 283, 315 (S.D.N.Y. 2025).

In any event, this presents “a classic fact dispute” inappropriate for resolution on a motion to dismiss. OB.47-49. Courts have allowed DMCA claims to proceed on “information-and-belief pleading” and “generally available information about AI algorithms” because plaintiffs generally “lack access to the granular details about how [defendants’ AI] technology works.” *Lehrman v. Lovo, Inc.*, 2025 WL 1902547, at \*19 (S.D.N.Y. July 10, 2025); *e.g.*, *N.Y.T.*, 777 F.Supp.3d at 315. The federal rules require no more. Fed. R. Civ. P. 8(a) (permitting pleading “on information and belief”); *Quintara Biosciences, Inc. v. Ruifeng Biztech, Inc.*, 2025 WL

2315671, at \*8 (9th Cir. Aug. 12, 2025) (trade secrets need not be alleged with particularity because that is a “fact question” to “be resolved on summary judgment or at trial”).

This Court has therefore permitted plaintiffs to establish CMI removal circumstantially by showing a “striking similarity” between the works.” *Friedman v. Live Nation Merch., Inc.*, 833 F.3d 1180, 1188 (9th Cir. 2016). When striking similarities exist and “CMI [is] missing,” courts may reasonably infer “the CMI had been removed on the copied version.” *Id.* at 1188-89 (“circumstantial evidence can be used to prove any fact”).<sup>6</sup> This is particularly true for AI models, where the relevant facts are uniquely in the defendant’s possession. 3-ER-221 ¶ 140;

---

<sup>6</sup> OpenAI tries to limit *Friedman* to “exact copies.” OAB.41. But *Friedman* explained that while “exact copies” created a “**compelling inference**” that CMI “**necessarily**” had been removed, a “striking similarity” may create a “**permissible inference**” of CMI removal. 833 F.3d at 1188 (emphasis added). GitHub’s attempt to dismiss *Friedman* as “an *infringement* claim,” GAB.52, ignores that the “striking similarity” analysis addressed section 1202(b) claims. 833 F.3d at 1187-88.

Several courts have adopted *Friedman*’s “striking similarities” inference for DMCA claims. *E.g.*, *Aardwolf Indus., LLC v. Abaco Machines USA, Inc.*, 2017 WL 10350547, at \*11 (C.D. Cal. Nov. 13, 2017) (considering “circumstantial evidence of access and subsequent publication of altered CMI”); *Izmo*, 2019 WL 13210561, at \*3; *Oracle Int’l Corp. v. Rimini St., Inc.*, 2023 WL 4706127, at \*82 (D. Nev. July 24, 2023).

*Lehrman*, 2025 WL 1902547, at \*19; *Izmo, Inc. v. Roadster, Inc.*, 2019 WL 13210561, at \*3 (N.D. Cal. Mar. 26, 2019) (at the pleading stage, “it is difficult and unfair to burden a [DMCA] plaintiff” with “produc[ing] specific facts” about “how the defendant came to possess the copyrighted materials”). That describes this case: Defendants had Plaintiffs’ code with CMI, and then their AI emitted strikingly similar code without CMI. That is certainly enough to plead a section 1202(b) violation.

2. Tellingly, Defendants seek to avoid addressing the merits of Plaintiffs’ allegations by misrepresenting that Plaintiffs raised the removal-from-training-data theory for the first time on appeal. OAB.54-57; GAB.64-66. Even if that were true, an output theory more than suffices to state a claim. But it is not true. The SAC includes the removal-from-training-data theory, alleging:

- Defendants “caused Codex and Copilot to ingest and distribute Licensed Materials,” 3-ER-195 ¶ 49;
- “Codex and Copilot have no way to determine whether license text or other [CMI] is part of the code it appears immediately before or after,” 3-ER-221 ¶¶ 141-42;
- “Early iterations of Copilot ... would sometimes produce” CMI with its outputs, 3-ER-222 ¶¶ 143-44;

- Defendants “strip[]” the code’s CMI “after they were uploaded to ... GitHub repositories” and **before** “distribut[ing] the now-anonymized code to Copilot users,” 3-ER-187, 235 ¶¶ 10-11, 215.

Removal before training is a plausible explanation for how Defendants stop Copilot from regurgitating CMI with its outputs. *See Ariz. Students’ Ass’n v. Ariz. Bd. of Regents*, 824 F.3d 858, 864 (9th Cir. 2016) (“accept the complaint’s well-pleaded factual allegations as true, and construe all inferences in the plaintiff’s favor”); *Skinner v. Switzer*, 562 U.S. 521, 530 (2011) (“complaint need not pin plaintiff’s claim for relief to a precise legal theory”). One solution is to prevent their AI models from memorizing CMI, by stripping CMI from *all* source code *before* the models memorize the training data. Plaintiffs preserved this theory by noting that “because Defendants control all the information about the training dataset, ... only Defendants know *when* the Licensed Materials ... were scraped.” 3-ER-221 ¶ 140.

Defendants complain Plaintiffs fail to specify the precise technical mechanism by which CMI was removed before training. GAB.66; OAB.56. Pleadings, however, require only “a short and plain statement,” not engineering specifications. Fed. R. Civ. P. 8(a)(2). The exact mechanism by which removal occurred is a factual question for discovery.

*See Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570, 556 (2007) (requiring only “enough fact to raise a reasonable expectation that discovery will reveal evidence”). “Plaintiffs need not prove the allegations ... at the motion to dismiss stage.” *Sonterra Cap. Master Fund Ltd. v. UBS AG*, 954 F.3d 529, 534 (2d Cir. 2020).

Plaintiffs also raised the training-data theory below. GAB.65-66; OAB.58. Plaintiffs argued before the district court that Defendants’ “removal of CMI from digital copies ... before distribution” violated section 1202(b), SER-14, which GitHub acknowledged below, 1-FER-28 (“Plaintiffs now focus on the pre-distribution removal of CMI.”). And Plaintiffs’ focus on Copilot’s outputs was responsive to Defendants’ categorical argument below that those outputs must be identical. *E.g.*, SER-20 (“Even if ‘identity’ was actually an element ... any differences ... are ‘immaterial.’”). The precise timing of removal (before or after training) was immaterial to the district court’s identity holding. SER-20-22. Only when Defendants signaled they were abandoning their categorical position on appeal did that timing become relevant—which is why Plaintiffs preserved the argument in their opening brief. 3-ER-138; OB.46-47.

Defendants thus performed the real “bait-and-switch” here. OAB.56. They made one argument below, completely changed their position on appeal, and now try to draw attention away from this maneuver by accusing Plaintiffs of somehow waiving the right to respond to Defendants’ brand-new argument. The Court should reject such gamesmanship. *See M.D. Russell Constr., Inc. v. Consol. Staffing, Inc.*, 2023 WL 8798086, at \*7 (4th Cir. Dec. 20, 2023) (party “cannot now make a completely contradictory argument on appeal”).

**B. Defendants’ Myopic Focus on Identicality as an Evidentiary Factor Contradicts Their Concessions and Fails on the Merits**

Defendants try to salvage their and the district court’s identicality requirement by recasting it as a “weighty factor” or “dispositive indication” that CMI was not removed from an existing work. GAB.50-51; OAB.50. This new version of an identicality standard is illogical and inconsistent with the statutory text and case law interpreting it.

1. Defendants’ logic is flawed. They argue that because identical copies missing CMI suggest removal, non-identical copies must mean *no* removal occurred. But this “confuses the sufficient with the necessary.” *United States v. Hernandez*, 37 F.4th 1316, 1320 (7th Cir. 2022); *In re*

*Ward*, 21 F.3d 1119, \*2 n.6 (9th Cir. 1994). Yes, “when a defendant possesses a work identical to the plaintiff’s work but without any CMI,” that suggests he removed CMI. OAB.22. But Defendants wrongly leap to the inverse proposition—that “lack of identicality ... confirms” no removal. OAB.43. Just because identical copies may indicate CMI removal does not mean non-identical copies ***disprove*** removal.

Consider an example. If a defendant steals a book, tears out the copyright-information page and the last chapter, and then sells the book, no one would doubt he removed CMI even though the book was also modified. Direct evidence—the tearing out—proves it. It would be “incongruous to require a plaintiff, in order to survive a motion to dismiss, to plead more facts than he [would] ultimately need to prove to succeed on the merits if direct evidence ... is discovered.” *Swierkiewicz*, 534 U.S. at 511-12. It thus cannot be the case that non-identicality is a “dispositive indication” no removal occurred.

Precedent confirms this point. In *GC2 Inc. v. Int’l Game Tech.*, non-identicality was irrelevant because there was direct evidence of CMI removal. 391 F.Supp.3d 828, 844 (N.D. Ill. 2019). The defendants received plaintiff’s digital artwork and published an edited version

without it. *Id.* *GC2* rejected the argument Defendants make here: that “because they modified the artwork ... creating a derivative work, there was no longer an ‘original work’ from which to remove [CMI].” *Id.* Likewise, Defendants cannot escape liability for CMI removal simply by programming their models to make trivial modifications before outputting. 3-ER-225 ¶ 155.

By contrast, *Friedman* needed circumstantial evidence because there was no direct proof defendants possessed plaintiff’s work. 833 F.3d at 1188. Thus, “striking similarities”—exact copies with unique “captions” and “layout features”—were necessary to show the pictures had been “scanned from [plaintiff’s] book.” *Id.* But here, the allegations show Defendants possessed Plaintiffs’ code and removed CMI. *Supra* pgs. 8-9, 11-12. Whether Defendants’ outputs are exact copies, modified versions, or verbatim excerpts is therefore immaterial.

Defendants’ reliance (OAB.38-39) on *Falkner v. General Motors*, 393 F.Supp.3d 927 (C.D. Cal. 2018), reinforces this point. The issue was not that defendant’s photograph was non-identical, but that capturing only the unsigned wall of a two-wall mural never created a copy with CMI to remove. *Id.* at 938. As OpenAI explains, however, downloading the

defendant’s photograph and deleting its metadata would remove CMI—regardless of what other modifications are made. OAB.35-36. The latter scenario mirrors this case. Defendants copied Plaintiffs’ complete works (CMI included) and programmatically stripped CMI from those works (either before training or before outputting)—classic DMCA violations.

2. More fundamentally, lack of identity cannot be dispositive evidence against removal because Defendants and their amici concede that superficial changes do not defeat liability. *E.g.*, IPAB.31-32. Defendants admit it would be “absurd” if Plaintiffs’ examples (at OB.38) could preclude liability, and they concede “trivial changes” can be made after removing CMI. GAB.50-51; OAB.46-47.<sup>7</sup> That describes Plaintiffs’ allegations, which identify several instances where Copilot reproduced “verbatim cop[ies]” of Plaintiffs’ code with only “immaterial” or “semantically insignificant” modifications—such as “cosmetic variations in line breaks.” 3-ER-208, 218-21 ¶¶ 108, 135-37. By Defendants’ own admission, these allegations state a claim.<sup>8</sup>

---

<sup>7</sup> OpenAI gives examples of how one could substantially modify a work yet still remove CMI—such as by defacing a painting “then strip[ping] the artist’s signature.” OAB.46.

<sup>8</sup> OpenAI wrongly claims “no case disputes that identity (or lack thereof) typically reflects whether” CMI was removed or altered.

3. Even so, Defendants insist that non-identity is generally “dispositive” because “in most cases,” GAB.46, “when someone strips CMI from a work, what remains is an identical copy,” OAB.41. This is only true in the most straightforward cases. *E.g.*, GAB.38 (offering “altering the title page of a book” as a prototypical DMCA violation). A typical violation is rarely so obvious. Rather, the “truly purposeful actors” targeted by section 1202(b)’s scienter requirement will usually act to conceal their violations. OB.33-34. Courts have seen this pattern repeatedly. For example, the defendants in *GC2* tried to evade liability by “modif[ying] the artwork.” 391 F.Supp.3d at 844.

Plaintiffs allege similar concealment here—that the “propensity for small cosmetic variations” is a “feature, not a bug,” and that Defendants intentionally program “Copilot to produce” such variations “as often as possible” to “conceal” reproductions and CMI removal. 3-ER-225 ¶ 155. An identity standard would allow bad actors to remove CMI with

---

OAB.48. *Contra Friedman*, 833 F.3d at 1188; *GC2*, 391 F.Supp.3d at 844; *supra* n.6 (listing cases accepting “striking similarities”); *compare* OAB.50 (suggesting *Oracle* merely “reject[ed] stringent identity requirement”), *with Oracle*, 2023 WL 4706127, at \*82 (endorsing liability where “modifie[d] source code [is] ‘substantially similar’ to Plaintiff’s copyrighted source code”).

impunity by making modifications afterwards—especially if non-identity forecloses discovery.

That would turn the statute on its head. The DMCA was designed to address “the ease with which software (and other electronic media) can be copied” and “distribute[d]” in “digital form.” *Microsoft Corp. v. AT&T Corp.*, 550 U.S. 437, 458 (2007). This concern is magnified for AI products that can systematically copy, modify, and redistribute vast amounts of digital works instantaneously. The DMCA becomes an empty letter if it cannot prevent sophisticated actors from using advanced technologies to circumvent the very protections Congress established for the digital age.

4. Ultimately, the degree of similarity between Copilot’s outputs and Plaintiffs’ code is a red herring. As Defendants acknowledge, “section 1202(b) is about making unauthorized changes to CMI on existing works,” “*not about making unauthorized copies of works, whether identical or not.*” OAB.46; GAB.53. This makes sense given the DMCA’s purpose “to combat copyright piracy in its earlier stages, **before** the work was even copied.” *Universal City Studios, Inc. v. Corley*, 273 F.3d 429, 435 (2d Cir. 2001) (emphasis added).

The proper inquiry remains what it has always been: Was CMI removed? Defendants and Plaintiffs agree a section 1202(b)(1) violation occurs the moment CMI is stripped from Plaintiffs' code. That is precisely what Plaintiffs allege. *Supra* pgs. 8-9, 11-12. Whatever else Defendants did to those works—whether they copied identically, made trivial or substantial modifications, or excerpted snippets—is irrelevant. OAB.46.

**C. Plaintiffs Do Not Ask for an Attribution Right Under the DMCA**

Plaintiffs have never suggested section 1202(b) requires adding CMI to new works. GAB.52; OAB.43-45. Plaintiffs have consistently maintained the statute prohibits only the removal of existing CMI, not the failure to add CMI to new creations.<sup>9</sup> Plaintiffs' only argument on appeal is that CMI *can* be removed from non-identical copies, OB.37-38, which even Defendants now concede, GAB.46, 51; OAB.46-47.

Defendants merely dispute what degree of similarity permits an inference of removal when direct evidence is unavailable—a familiar inquiry in copyright law. *See* OB.26, 42. But the Court need not answer that question here because Plaintiffs allege direct evidence of removal:

---

<sup>9</sup> Thus, amici's arguments and "public policy perspective[s]" are irrelevant to this appeal. AAB.4, 10-13; CPAB.5-18; PKAB.8-14.

Defendants had identical copies of Plaintiffs' source code with CMI and then removed it. *E.g.*, 3-ER-213, 221-22 ¶ 120, 141-42.

OpenAI asserts “no case” “expressly rejects” its position that non-identity suggests a new work was created. OAB.48-50. But multiple courts have done just that. *New Parent World, LLC v. True To Life Prods. Inc.*, squarely rejected the argument that non-“substantially similar” copies indicate a “distinct work.” 2024 WL 4277865 (D. Ariz. Sept. 24, 2024). *ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.*, disagreed that “superficial alterations” indicate a “new work.” 667 F.Supp.3d 411, 428 (S.D. Tex. 2023). And *Real World Media, LLC v. Daily Caller, Inc.*, held “nothing in § 1202(b) requires precise equivalence,” finding such “work-specific arguments” better suited for summary judgment. 744 F.Supp.3d 24, 40 & n.8 (D.D.C. 2024).

None of these decisions somehow transformed copyright law by rejecting Defendants' position. Defendants and their amici warn that unless this Court adopts their new theory, “merely” creating a “derivative” or substantially similar work would violate the DMCA. GAB.52; OAB.34; AAB.8; CPAB.9-10. But as Defendants admit, whether

a copy is identical, infringing, or derivative is irrelevant. OAB.46. All that matters is whether CMI was removed with scienter.

Indeed, simply “recast[ing],” “adapt[ing],” or “lift[ing] some elements” from “preexisting works” proves nothing about CMI removal. 17 U.S.C. § 101; GAB.52. Thus, creating derivative works alone does not violate the DMCA. Artists who create “substantially similar” works face no DMCA liability for omitting attribution—“Blurred Lines” and Warhol’s “Prince Series” remain safe. OAB.52.

But the converse is equally true. Defendants cannot escape DMCA liability for removing CMI from digital copies simply by also modifying the work, even if that creates a derivative (or substantially similar) work. Take GitHub’s *Twilight* fan fiction example. GAB.53-54. Writing vampire romance inspired by that series creates no DMCA liability. But if someone received Stephanie Meyer’s manuscript, deleted the title page, and “transformed” it into a derivative work, they would still be liable for removal. What matters is that CMI was removed.

The distinction matters here. This case is not about artists drawing inspiration or musicians unconsciously borrowing melodies. OAB.52. It is not even about coders studying the work of peers and applying lessons

from it. Instead, Defendants downloaded billions of lines of licensed code, programmatically stripped their CMI, and built a commercial computer system to redistribute that CMI-stripped code to millions of consumers. OB.9-13, 50-53. Regardless of whether Copilot outputs “verbatim copies” or “near-identical copies,” 3-ER-259, Defendants industrialized CMI removal at an unprecedented scale—precisely the technological threat Congress anticipated with the DMCA. *Corley*, 273 F.3d at 435.

### **III. PLAINTIFFS HAVE STANDING**

The district court found Plaintiffs have standing. 3-ER-252-53. It was correct, and Defendants’ standing arguments fare no better than their merits arguments. “At the pleading stage, general factual allegations of injury resulting from the defendant’s conduct may suffice” because courts must “presum[e] that general allegations embrace those specific facts that are necessary to support the claim.” *Lujan v. Defs. of Wildlife*, 504 U.S. 555, 561 (1992). Plaintiffs “need only show that the facts alleged, if proved, would confer standing upon” them. *Warren v. Fox Family Worldwide, Inc.*, 328 F.3d 1136, 1140 (9th Cir. 2003).

#### **A. Plaintiffs’ Injuries Are Concrete and Particularized**

Plaintiffs’ harms are neither “conjectural” nor “hypothetical.” OAB.24-25; GAB.27-29. Defendants’ contention that Plaintiffs allege

only a probability of harm misapprehends the allegations, precedent, and statistics.

**1. *CMI Removal Is an Existing, Concrete Injury***

Plaintiffs’ intellectual property interests were harmed by removal of their CMI, before any distribution occurred. Section 1202(b)(1) prohibits “intentionally removing or altering” CMI. Plaintiffs allege a concrete, particularized injury that occurred the moment Defendants stripped CMI. Whether this occurred before feeding the data into the models or after training, the result is the same—Defendants removed CMI from Plaintiffs’ works.

This injury is neither “conjectural” nor “hypothetical.” Because the violation occurred at removal, the injury already exists. The harm is not just the likelihood of future outputs without CMI; it also results from the completed act of removing CMI from Plaintiffs’ code, irrespective of Defendants’ subsequent distribution. *Cf. Friedman*, 833 F.3d at 1187 (court erred by “focus[ing]” only on section 1202(b)(1) and ignoring (b)(3)).

Defendants’ removal constitutes a concrete injury to Plaintiffs’ property interests in their licensed works. As *TransUnion LLC v. Ramirez* recognized, “traditional tangible harms, such as physical harms

and monetary harms,” are the “most obvious” injuries sufficient for Article III standing. 594 U.S. 413, 425 (2021) (“reputational harms” are also concrete). Unauthorized CMI removal injures intellectual property rights, which common law and Congress have long recognized, *Intercept*, 767 F.Supp.3d at 26-28, and which Congress further codified in the DMCA. When Congress creates a statutory prohibition, courts “must afford due respect to Congress’s decision.” *TransUnion*, 594 U.S. at 425.

The removal or alteration of CMI frustrates the DMCA’s core purpose—“to protect consumers from misinformation as well as authors and copyright owners from interference with the private licensing process.” *Fischer v. Forrest*, 968 F.3d 216, 222 (2d Cir. 2020); 3-ER-201 ¶ 76. Section 1202(b) was designed to ensure the “[i]ntegrity of copyright management information,” which is critical for “tracking and monitoring uses of copyrighted works, as well as licensing of rights and indicating attribution, creation and ownership.” S. Rep. No. 105-190 (1998) at 16.

Accordingly, Defendants’ unauthorized CMI removal is “an interference with a property right that is ‘closely related to the kind of property-based harms traditionally actionable in copyright.’” *Kadrey*, 2025 WL 744032, at \*1. Under Title 17, section 1202(b)(1)’s prohibition

against CMI removal is just like section 106(1)'s exclusive right of reproduction: both describe a concrete harm to intellectual property before distribution, itself a distinct cause of action under sections 1202(b)(3) and 106(3).

Here, Plaintiffs allege Defendants' removal "conceals the copying of Licensed Materials with the intent and purpose of making it difficult for Plaintiffs and members of the class to identify breaches of the licenses and to enforce their rights." 3-ER-225 ¶ 155; 3-ER-241 ¶ 257. That matters because open-source licenses provide economic benefits through attribution requirements that increase the developers' reputations and enable them to monitor improvements to their projects.<sup>10</sup> 3-ER-201 ¶ 76; *Jacobsen v. Katzer*, 535 F.3d 1373, 1378-79 (Fed. Cir. 2008). When Defendants stripped CMI, they denied Plaintiffs the benefit of these property interests—a concrete harm under Article III.

Defendants argue Plaintiffs lack standing because they cannot show their specific code will be distributed by Defendants' products in the future. OAB.28; GAB.30. But, as Judge Rakoff recently explained,

---

<sup>10</sup> That Plaintiffs' code is open source does not make them available for unrestricted use. Plaintiffs chose to restrict their use by requiring CMI in their licenses. 3-ER-201, 214 ¶¶ 76, 124.

section 1202(b) claims need not “identify a public dissemination” to establish a “concrete injury” because the traditional “injury to a property right”—including “copyright injury”—“does not require publication to a third party.” *Intercept*, 767 F.Supp.3d at 28-29. Just like if a painting was recklessly damaged, one need not show or sell the damaged painting to injure the artist’s property rights.

## **2. *These Injuries Are Redressable, Not Moot***

OpenAI’s redressability and mootness arguments mischaracterize both Plaintiffs’ injuries and the available relief. OAB.30-32.

First, these arguments ignore Plaintiffs’ claim under section 1202(b)(3), which Defendants violate each time they knowingly “distribute” Plaintiffs’ code without CMI. *See Friedman*, 833 F.3d at 1187 (noting plaintiff did not have to prove intentional removal because he could still “prevail by showing distribut[ion]” under section 1202(b)(3)).

Second, the requested injunction—requiring CMI inclusion whenever Defendants reproduce Plaintiffs’ code—would redress the “ongoing” and “future injury” caused by Defendants’ removal. *Overbey v. Mayor of Baltimore*, 930 F.3d 215, 230 (4th Cir. 2019); 3-ER-238-39 ¶ 233. Every time Copilot reproduces Plaintiffs’ code without attribution, it

perpetuates and compounds the original harm—the deprivation of the authors’ right to have CMI accompany their work, which protects their economic and reputational interests. *Fischer*, 968 F.3d at 222. Disseminating Plaintiffs’ code without “statutorily guaranteed” CMI harms their “concrete interest” in the availability and accuracy of that information. *Cf. Magadia v. Wal-Mart Assocs., Inc.*, 999 F.3d 668, 679 (9th Cir. 2021); *Robins v. Spokeo, Inc.*, 867 F.3d 1108, 1113 (9th Cir. 2017) (“dissemination of false information in consumer reports” is a “concrete harm” because “threat to a consumer’s livelihood is caused by the very existence of inaccurate information”). Thus, Plaintiffs’ injuries did not end with the initial removal of CMI, OAB.31, but are continuing.<sup>11</sup>

Indeed, courts routinely enjoin ongoing harm from past violations. *E.g., Davidson v. Kimberly-Clark Corp.*, 889 F.3d 956, 972 (9th Cir. 2018) (inability to trust “flushable” wipe labels due to past “false advertising” is “certainly impending” injury redressable by injunction requiring “truthful representations”); *Mayfield v. United States*, 599 F.3d 964, 971-

---

<sup>11</sup> Plaintiffs are also injured by CMI removal resulting from ongoing training—either from new code uploaded to GitHub or on copies utilized by other GitHub members (consistent with the licenses). 3-ER-198, 221 ¶¶ 61 n.7, 140.

72 (9th Cir. 2010) (injunction would redress “present, on-going injury” from “government’s continued retention” of unlawfully seized material); *Overbey*, 930 F.3d at 230. Congress specifically authorized injunctive relief “to prevent or restrain” section 1202(b) violations precisely because monetary damages cannot redress continuing distribution. 17 U.S.C. § 1203(b)(1); 3-ER-238-39 ¶ 233. Without an injunction, Copilot will continue disseminating Plaintiffs’ code without attribution—“precisely the sort of ‘real harm’” section 1202(b) “was designed to prevent.” *Magadia*, 999 F.3d at 680.

Because an injunction would provide effectual relief, Plaintiffs’ training-data theory is not moot. *See Knox v. Serv. Emps. Int’l Union*, 567 U.S. 298, 307 (2012) (case is “moot only when it is impossible” to “grant any effectual relief.”); *Malik v. DHS*, 78 F.4th 191, 199 (5th Cir. Aug. 15, 2023) (injunction not moot where defendant “still has the [plaintiff’s] data”).

### **3. *Plaintiffs Have Shown Substantial Risk of Future Harm by Concrete Evidence, Not Speculation***

With respect to Plaintiffs’ section 1202(b)(3) claims, Defendants mischaracterize Plaintiffs’ showing as merely a “probabilistic estimate” of future harm. GAB.26-29; OAB.25-26. But Defendants’ own

admissions, empirical data, and concrete examples show a virtual certainty of ongoing injury that easily satisfies Article III.

“An allegation of future injury” establishes standing if “there is a substantial risk that the harm will occur.” *Susan B. Anthony List v. Driehaus*, 573 U.S. 149, 158 (2014) (quotations omitted). The evidence here demonstrates a near-certainty that Defendants’ AI product will continue reproducing Plaintiffs’ code without CMI. At least three mutually reinforcing data points confirm this. First, GitHub’s own internal research shows that Copilot’s outputs match publicly available code “about 1 percent of the time.” 3-ER-206 ¶ 102. Second, studies show that the rate of verbatim reproductions is increasing over time. 3-ER-207 ¶ 104. Third, Plaintiffs have demonstrated Copilot *can* and *has* produced their code without CMI. 3-ER-206-21 ¶¶ 99-139.

Contrary to Defendants’ suggestions, OAB.26; GAB.26-29, statistical evidence can support a substantial risk of future harm, which is inherently probabilistic. *See Food & Water Watch, Inc. v. Vilsack*, 808 F.3d 905, 918 n.6 (D.C. Cir. 2015) (plaintiffs may “rel[y] on statistics to show a substantial increase in the risk of harm”). Courts have found substantial risk in analogous data-breach cases based on similar

statistical likelihoods. *See Galaria v. Nationwide Mut. Ins. Co.*, 663 F. App'x 384, 386 (6th Cir. 2016) (“substantial risk” where study showed data-breach victims were “9.6 times more likely to experience identity fraud”); *In re Zappos.com, Inc.*, 888 F.3d 1020, 1027 (9th Cir. 2018) (same where “information taken in the breach” placed plaintiffs “at higher risk of ... identify theft”). Like data-privacy laws, Congress designed section 1202(b) as a prophylactic measure to reduce the risk of future infringement. *See id.* (credit cards “warrant legislation specifically to reduce the risk of identity theft”). Just as “substantial risk” exists where plaintiffs’ “data has already been stolen and is now in the hands of ill-intentioned” parties, *Galaria*, 663 F. App'x at 388, substantial risk exists here where Defendants already removed CMI to conceal infringement, 3-ER-225 ¶ 155. *Cf. Lewert v. P.F. Chang's China Bistro, Inc.*, 819 F.3d 963, 966 (7th Cir. 2016) (where “data theft” “*already* occurred,” plaintiffs “should not have to wait until hackers commit” fraud).

*Summers v. Earth Island Inst.*, 555 U.S. 488 (2009), does not prohibit statistical evidence of likely future harm. *Summers* addressed a fundamentally different scenario: “organizational standing,” where the organization could not identify a single affected member. *Id.* at 497-99.

In that context, the Supreme Court held that “naming the affected members has never been dispensed with in light of statistical probabilities.” *Id.* Without showing particularized harm to even one member, it was insufficient to merely assert that some unidentified members would be injured somewhere, somehow, sometime in the future. *Id.*; e.g., *FDA v. All. for Hippocratic Med.*, 602 U.S. 367, 382 (2024); *Satanic Temple v. Labrador*, 2025 WL 2302188, at \*3 (9th Cir. Aug. 11, 2025) (“absence” of any identified injured member, “coupled with infirm statistical inference, does not meet Article III’s standing requirements”).

But “unlike *Summers*, this is not a case resting on unsupported ‘statistical probability’ for organizational standing.” *Massachusetts v. HHS*, 923 F.3d 209, 223 (1st Cir. 2019). Plaintiffs seek relief for themselves based on personalized injuries, not injuries to unidentified parties.<sup>12</sup> Plaintiffs’ statistical evidence is combined with independent proof that they are “among the injured,” *Lujan*, 504 U.S. at 563, and

---

<sup>12</sup> Accordingly, this case does not implicate concerns about third-party or generalized harms that animate *Summers* and its progeny. E.g., *All. for Hippocratic Med.*, 602 U.S. at 399-405 (Thomas, J., concurring); *Doe v. Hochul*, 139 F.4th 165, 182 (2d Cir. 2025) (no standing based on statistical likelihood that “at some point in the future, an unidentified woman may seek to obtain an abortion of an unidentified fetus from an unidentified abortion provider”).

suffered injuries “in a personal and individual way,” *Spokeo, Inc. v. Robins*, 578 U.S. 330, 339 (2016). Plaintiffs show Defendants already injured them. And “one way to help show imminent future harm is to show past exposure to the same kind of harm from the same defendant.” *Tennessee Conf. of NAACP v. Lee*, 139 F.4th 557, 567 (6th Cir. 2025).

Plaintiffs therefore allege far more than a mere “statistical probability.” OAB.26; GAB.29. This Court has explained that standing based on a threat of future harm exists when some systematic pattern repetition, or relationship exists.” *Nelsen v. King County*, 895 F.2d 1248, 1254 (9th Cir. 1990). Here, the combined showing of (1) confirmed inclusion of Plaintiffs’ code in training data, (2) systematic CMI removal on an industrial scale; (3) Defendants’ admission of regular code reproduction, (4) hundreds of millions of daily queries, and (5) prior incidents of harm to Plaintiffs specifically, is far more than a “naked statistical assertion” based on “general circumstances.” *Id.* at 1250-51.

This is not a case where plaintiffs allege only “probabilistic standing based on a non-particularized risk” that “equally affects the general public.” *E.T. v. Paxton*, 41 F.4th 709, 715 (5th Cir. 2022). Plaintiffs have established a systemic pattern of repetition and an ongoing relationship

with Defendants, who *continue* to use and reproduce Plaintiffs' code. *See Mayfield*, 599 F.3d at 970-71 (“injury is likely to recur” where “harm is part of a pattern of officially sanctioned” conduct).

Defendants' cases involve materially different circumstances. *Raw Story Media, Inc. v. OpenAI, Inc.*, involved an AI model “trained on a ‘scrape of most of the internet’” with no examples of reproduction or allegations about frequency or likelihood. 756 F.Supp.3d 1, 7 (S.D.N.Y. 2024). Here, Defendants' models were trained on a targeted dataset, Copilot already reproduced Plaintiffs' code without CMI, and statistical evidence shows the frequency of future reproductions.

Plaintiffs' injuries do not depend on a “chain of highly speculative contingencies.” *Nelsen*, 895 F.2d at 1250. In *Nelsen*, future injury required eight independent contingencies: that plaintiff “remain within [the c]ounty, remain indigent,” relapse into alcoholism, “commit an alcohol-related offense, be prosecuted” and “convicted, be offered the choice to reenter the [rehab] Center, make that choice, *and* find” unchanged conditions. *Id.* Similarly, to suffer injury, the plaintiff in *Los Angeles v. Lyons* would have had to encounter police, engage in arrest-worthy behavior, and experience the same unlawful response from law

enforcement. 461 U.S. 95, 105-06 (1983). And *Pinkert v. Schwab Charitable Fund* depended entirely on plaintiff's hypothetical future donation decisions. 48 F.4th 1051, 1055 (9th Cir. 2022) (plaintiff did not allege "concrete plans to donate particular amounts in the future" or past pattern). These attenuated sequences were too speculative.

Here, no such conjecture is required. Defendants already copied Plaintiffs' code, removed CMI, and distributed verbatim excerpts without CMI. Defendants admit their AI frequently reproduces CMI-less copies, and studies show increasing reproduction rates. These allegations do not rely on multiple future, hypothetical decisions by Plaintiffs or others.

Additionally, Defendants' attempts to trivialize the one-percent figure misapprehend the scale. It need "not be 'literally certain' that Plaintiffs' [code] will be misused" again. *Galaria*, 663 F. App'x at 388. With Copilot now processing over 400 million queries a day<sup>13</sup> from over 15 million users,<sup>14</sup> even "a mere 1%" represents more than 1.4 billion matches a year, based on Defendants' own data. This figure continues to

---

<sup>13</sup> <https://www.infoq.com/presentations/github-copilot/>.

<sup>14</sup> Microsoft now reports "over 15 million GitHub Copilot users, up over 4X year-over-year." <https://www.microsoft.com/en-us/investor/events/fy-2025/earnings-fy-2025-q3>.

climb as the rate of reproductions increases along with Copilot's user base. 3-ER-207 ¶¶ 104-05. Given enough time, this represents at least a substantial risk that Plaintiffs' code will continue to be among the billions of reproductions emitted every year.

GitHub argues Plaintiffs have not alleged their "code appears frequently enough to be part of any detectable statistical pattern." GAB.28; OAB.27-28. This misstates how large language models work. Copilot does not just reproduce frequently appearing code; it memorizes and regurgitates training data, with memorization increasing as the "model's capacity grows." 3-ER-207-08 ¶¶ 104-05. The duplicate-detection feature confirms Copilot regularly produces exact matches. 3-ER-222-23 ¶¶ 144-49. While OpenAI claims this feature makes exact matches "even less likely," OAB.26, it is optional, 3-ER-223 ¶ 148, and Defendants program Copilot to make "cosmetic modifications to conceal" reproduction, 3-ER-225 ¶ 155. At most, this is a merits or damages question. *See Intercept*, 767 F.Supp.3d at 31 (explaining plaintiffs need only allege AI product "has some capacity to regurgitate," not that plaintiffs' regurgitations are "similar to all such regurgitations").

## B. Plaintiffs' Injuries Are Not Manufactured

Next, Defendants accuse Plaintiffs of “manufactur[ing]” their injuries by prompting Copilot with their own code. OAB.28-30; GAB.30-32. But the examples of Copilot reproducing Plaintiffs' code are not attempts to manufacture standing; they are evidence of an injury that already occurred.

1. Article III distinguishes between creating artificial injuries and exposing existing violations through deliberate testing. Courts have long recognized plaintiffs may establish standing by testing whether defendants are complying with federal law. In *FEC v. Cruz*, the Supreme Court squarely rejected that plaintiffs “lack standing” for “self-inflicted” injuries when they “knowingly triggered” the challenged campaign finance rule. 596 U.S. 289, 296 (2022). Injuries “resulting from the application or threatened application of an unlawful” rule remain “fairly traceable” to defendants, “even if the injury could be described in some sense as willingly incurred.” *Id.* at 297. What matters is whether defendant's unlawful conduct caused the harm, not whether plaintiffs deliberately exposed themselves to it. *Id.* (citing *Havens Realty Corp. v. Coleman*, 455 U.S. 363, 374 (1982) (holding tester posing as renter did

not manufacture injury by subjecting himself to discriminatory practices that were already occurring)).

Plaintiffs’ testing merely demonstrated ongoing conduct by Defendants’ AI product. The testing examples provide concrete proof of what Defendants’ own internal data already confirmed: that Copilot regularly reproduces training data. 3-ER-222-23 ¶¶ 144-49. When Copilot can be prompted to produce verbatim copies of Plaintiffs’ code without CMI, it demonstrates Defendants possess and use copies of Plaintiffs’ works with CMI stripped. This is analogous to a plaintiff in a data breach case running a credit check to confirm his information was compromised. The check does not create the injury—it confirms it.

This case is nothing like *Clapper v. Amnesty Int’l USA*, 568 U.S. 398 (2013). GAB.31; OAB.28. There, plaintiffs attempted to “manufacture standing by choosing to make expenditures based on” speculative fears of future surveillance that might never occur. 568 U.S. at 402, 416. The plaintiffs had no greater risk of being surveilled than any other member of the public and “fail[ed] to offer any evidence that their communications” ever had been, were currently, or ever would be

“monitored.” *Id.* at 411 (plaintiffs lacked “knowledge of the Government’s § 1881a targeting practices”).

But there is nothing speculative about Defendants’ conduct. All Plaintiffs’ testing did was reveal that Copilot reproduces their code without CMI—a capacity that exists independent of any testing. Unlike *Clapper*, Plaintiffs did not manufacture a hypothetical injury; they exposed ongoing harm through targeted investigation. *See Cruz*, 596 U.S. at 297 (distinguishing *Clapper* on similar grounds). Using short snippets of their code as prompts does not “manufacture” the injury any more than a litigant “manufactures” a First Amendment violation by speaking. The injury is caused by Copilot’s design and operation, not in Plaintiffs’ method of revealing it. Indeed, there is an easy way for Defendants to avoid inflicting injury from testing: do not violate the law.

2. Defendants argue Plaintiffs must show “that ordinary users are likely to enter prompts that will cause Codex or Copilot to generate *Plaintiffs’* code.” OAB.26-27; GAB.30. This is legally irrelevant. Even if Plaintiffs “chose to subject themselves” to Defendants’ conduct, Plaintiffs’ “injuries are directly inflicted by” Defendants’ actions. *Cruz*, 596 U.S. at 297. That is all Article III requires.

Moreover, this improperly “conflate[s] the concreteness injury with [Defendants’] arguments on the merits.” *Intercept*, 767 F.Supp.3d at 29. For standing, courts must “accept as valid the merits of appellees’ legal claims.” *Cruz*, 596 U.S. at 298. In *Intercept*, Judge Rakoff rejected the same argument that plaintiffs lacked standing because they “cannot plausibly allege that any ChatGPT user has ever used the kind of prompt [plaintiff] used to generate similar regurgitated output.” 767 F.Supp.3d at 29 (finding standing based on similar demonstrations). Whether Plaintiffs’ demonstrations reflect the prompts of a “real-world user,” GAB.36, is a factual question requiring discovery into actual usage patterns—not a basis for dismissing on standing grounds.

Moreover, Defendants’ premise is wrong. Programmers routinely input existing functions as “prompts” for Copilot to complete—that is what Copilot is marketed to do. 3-ER-197 ¶¶ 58-59.

#### **IV. PLAINTIFFS ADEQUATELY ALLEGE SCIENTER**

Defendants invoke *Stevens v. CoreLogic*, 899 F.3d 666 (9th Cir. 2018), as a last-ditch effort to save the decision below. GAB.55-64. This argument should be rejected.

1. The district court never addressed *Stevens* in its dismissal order. 1-ER-8. This Court is “particularly” reluctant to decide fact-intensive issues—like knowledge or intent—that were not decided below.<sup>15</sup> *DeFries*, 104 F.4th at 1109.

2. *Stevens* does not apply to pleadings. GitHub misapplies *Stevens* by demanding that Plaintiffs prove “an objective likelihood of facilitating or concealing infringement.” GAB.57-58. To be sure, *Stevens* requires “an affirmative showing” that the defendants possessed the requisite mental state, “such as by demonstrating a past ‘pattern of conduct’ or ‘modus operandi.’” 899 F.3d at 674-75. But *Stevens*’ “affirmative showing” requirement explicitly refers to the “evidence” required at summary judgment, not pleading. *Id.* at 673, 675; *cf. Swierkiewicz*, 534 U.S. at 510. *Stevens* did not create a heightened pleading standard, and no circuit court has extended *Stevens* to require an “affirmative showing” of scienter on a motion to dismiss. *E.g., Bain v. Film Indep., Inc.*, 2022 WL 17592422, at \*1 (9th Cir. Dec. 13, 2022); *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 172-73 (2d Cir. 2020); *Zuma*

---

<sup>15</sup> The district court already found scienter requirement satisfied in the original complaint. GH-SER-135-36. That order is not on appeal, and Plaintiffs have since strengthened their allegations. GAB.62-63.

*Press, Inc. v. Getty Images, Inc.*, 845 F. App'x 54, 57-58 (2d Cir. 2021); *Victor Elias Photography, LLC v. Ice Portal, Inc.*, 43 F.4th 1313, 1320 (11th Cir. 2022); *Loeb-Defever v. Mako, LLC*, 2023 WL 5611042, at \*5 (5th Cir. Aug. 30, 2023).<sup>16</sup>

Mental states like “knowledge” and “intent” “may be alleged generally.” Fed. R. Civ. P. 9(b). Knowledge is a “question of fact” typically “inferred from circumstantial evidence,” since “direct proof” of a party’s “state of mind” is “rarely available” in section 1202(b) claims. *Friedman*, 833 F.3d at 1189. “[W]hether a party knew or should have known of a particular condition” is “inappropriate for resolution by summary judgment,” let alone dismissal. *F.T.C. v. Network Servs. Depot, Inc.*, 617 F.3d 1127, 1139 (9th Cir. 2010).

Thus, section 1202(b) scienter may be alleged generally, and knowledge may be inferred through circumstantial evidence. *See APL Microscopic, LLC v. Steenblock*, 2022 WL 4830687, at \*2 (9th Cir. Oct. 3,

---

<sup>16</sup> In GitHub’s three district court cases applying *Stevens* at the pleading stage, GAB.58-59 & n.6, plaintiffs did “not allege any facts” supporting scienter. *Harrington v. Pinterest, Inc.*, 2021 WL 4033031, at \*6 (N.D. Cal. Sept. 3, 2021); *Philpot v. Alternet Media, Inc.*, 2018 WL 6267876, at \*5 (N.D. Cal. Nov. 30, 2018). *Tremblay v. OpenAI, Inc.*, even contrasted those conclusory allegations against Plaintiffs’ concrete allegations here. 716 F.Supp.3d 772, 779 (N.D. Cal. 2024).

2022) (cropped screenshots posted on social media raised “reasonable inference” that defendant “knowingly removed” CMI); *Brown v. Stroud*, 2011 WL 2600661, at \*6 (N.D. Cal. June 30, 2011) (“knowledge and intent are better addressed” at “summary judgment”); *Izmo*, 2019 WL 13210561, at \*4.

3. Even if *Stevens* applied, Plaintiffs easily satisfy its requirements by showing a “pattern of conduct” or “modus operandi.” 899 F.3d at 675. Defendants’ systematic violations and admissions demonstrate their knowledge and intent:

- **Pattern of CMI Removal:** Defendants programmed their models to strip CMI before distribution—knowing this would produce verbatim excerpts without attribution. 3-ER-205-07, 222 ¶¶ 94-102, 143-44.
- **Defendants’ M.O.:** Defendants program Copilot to “produce small cosmetic variations” to “conceal the copying of Licensed Materials” and make “it difficult ... to identify breaches” of licenses. 3-ER-224-25 ¶¶ 145-47, 155.
- **Knowledge of Future Impact:** Defendants know Copilot still emits verbatim excerpts without CMI and could prevent this by not stripping CMI or blocking flagged code, but they choose not to—letting unwitting users infringe on protected code instead. 3-ER-199, 206-07, 222-25, 237 ¶¶ 68, 99-102, 145-155, 226.
- **Continued Violations Despite Notice:** Three years into this litigation, Defendants continue their practice despite knowing it facilitates and conceals infringement.

This is far more than a “general possibility” that removal will lead to infringement. *Stevens*, 899 F.3d at 673. This “pattern of conduct” shows Defendants are “aware” of the “probable future impact” of their actions. *Id.* at 675. By design, Defendants’ business model induces, enables, and conceals infringement—satisfying section 1202(b)’s scienter requirement. *See Kadrey*, 2025 WL 744032, at \*2 (scienter plausibly alleged where defendant “knew” AI model was “prone to memorizing and generating” CMI unless “removed from its training data” and took “steps to reduce th[at] likelihood”).

4. Finally, GitHub claims section 1202(b) requires proving actual infringement. GAB.59-62. This appears nowhere in the statute or *Stevens*. *Stevens* instead confirmed plaintiffs “need not show that any specific infringement has already occurred,” even at summary judgment. 899 F.3d at 674. Section 1202(b) uses future tense: removal that “*will*” induce, facilitate, or conceal infringement. *See Harrington*, 2021 WL 4033031, at \*6 (requiring only awareness of “probable future impact”).

GitHub’s proposed test—requiring plaintiffs to prove copyrightability, substantial similarity, and absence of licenses, and to overcome fair use defenses—would merge section 1202(b) with

infringement claims. This would contradict the statutory language and function: to prevent conduct that *facilitates* future infringement, not litigating completed infringement. GitHub's interpretation renders section 1202(b) superfluous.

### **CONCLUSION**

The Court should reverse the district court's dismissal of Plaintiffs' claims under §§ 1202(b)(1) and (b)(3) of the DMCA.

Dated: September 2, 2025

Respectfully submitted,

Jesse Panuccio  
Evan Ezray  
Daniel Morales  
BOIES SCHILLER FLEXNER LLP  
401 East Las Olas Blvd., Ste 1200  
Fort Lauderdale, FL 33301  
(954) 356-0011  
jpanuccio@bsfllp.com  
eezray@bsfllp.com  
dmorales@bsfllp.com

Maxwell V. Pritt  
Joshua M. Stein  
Margaux Poueymirou  
BOIES SCHILLER FLEXNER LLP  
44 Montgomery Street, 41st Floor  
San Francisco, CA 94104  
(415) 293-6800  
mpritt@bsfllp.com  
jstein@bsfllp.com  
mpoueymirou@bsfllp.com

s/ Joseph R. Saveri  
Joseph R. Saveri  
Christopher K.L. Young  
Evan Creutz  
William W. Castillo Guardado  
JOSEPH SAVERI LAW FIRM, LLP  
601 California Street, Suite 1505  
San Francisco, CA 94108  
(415) 500-6800  
jsaveri@saverilawfirm.com  
cyoung@saverilawfirm.com  
ecreutz@saverilawfirm.com  
wcastillo@saverilawfirm.com

Matthew Butterick  
(State Bar No. 250953)  
1920 Hillhurst Avenue, #406  
Los Angeles, CA 90027  
(323) 968-2632  
mb@buttericklaw.com

*Counsel for Plaintiff-Appellants and Proposed Class*

**UNITED STATES COURT OF APPEALS  
FOR THE NINTH CIRCUIT**

**Form 8. Certificate of Compliance for Briefs**

**9th Cir. Case Number(s):** 24-7700

I am the attorney or self-represented party.

This brief contains 8384 words, including 0 words manually counted in any visual images, and excluding the items exempted by FRAP 32(f). The brief's type size and typeface comply with FRAP 32(a)(5) and (6).

I certify that this brief (*select only one*):

complies with the word limit of Cir. R. 32-1.

is a **cross-appeal** brief and complies with the word limit of Cir. R. 28.1-1.

is an **amicus** brief and complies with the word limit of FRAP 29(a)(5), Cir. R. 29-2(c)(2), or Cir. R. 29-2(c)(3).

is for a **death penalty** case and complies with the word limit of Cir. R. 32-4.

complies with the longer length limit permitted by Cir. R. 32-2(b) because (*select only one*):

it is a joint brief submitted by separately represented parties.

a party or parties are filing a single brief in response to multiple briefs.

a party or parties are filing a single brief in response to a longer joint brief.

complies with the length limit designated by court order dated \_\_\_\_\_.

is accompanied by a motion to file a longer brief pursuant to Cir. R. 32-2(a).

Signature: s/ Joseph R. Saveri

Date: 9/2/2025

## CERTIFICATE OF SERVICE

I hereby certify that on September 2, 2025, I electronically filed the foregoing brief with the Clerk of the Court for the United States Court of Appeals for the Ninth Circuit using the Appellate CM/ECF system.

*s/ Joseph R. Saveri*  
Joseph R. Saveri

Dated: September 2, 2025