Case No. _____

United States Court of Appeals

for the

Ninth Circuit

_____

J. DOE 1, et al.,

Plaintiffs and Petitioners

v.

GITHUB, INC., et al.,

Defendants and Respondents

_____

From a Certified Order of the United States District Court
Northern District of California, Oakland Division
The Honorable Jon S. Tigar, Presiding
Case No. 4:22-cv-06823-JST

**PETITION FOR PERMISSION TO APPEAL ORDER
CERTIFIED UNDER 28 U.S.C. § 1292(b)**

JOSEPH R. SAVERI
**JOSEPH SAVERI LAW
FIRM, LLP**
601 California St., 1505
San Francisco, CA 94108
Telephone: (415) 500-6800
Email:
jsaveri@saverilawfirm.com

MATTHEW BUTTERICK
(State Bar No. 250953)
1920 Hillhurst Avenue,
#406
Los Angeles, CA 90027
Telephone: (323) 968-2632
Email:
mb@butticklaw.com

MAXWELL V. PRITT
JESSE PANUCCIO
**BOIES SCHILLER
FLEXNER LLP**
44 Montgomery St., 41st FL
San Francisco, CA 94104
Telephone: (415) 293-6800
Email: mpritt@bsfllp.com
jpanuccio@bsfllp.com

Attorneys for Plaintiff-Petitioners and the Proposed Class

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page(s)

Cases

Federal Statutes

I. INTRODUCTION

This appeal presents an unresolved and critical question of statutory interpretation under the Digital Millennium Copyright Act ("DMCA"): whether 17 U.S.C. § 1202(b) requires that copies of works be "identical" in order for liability to attach. The district court imputed such a requirement into the statute, even though it is found nowhere in the statutory text (nor ever mentioned in its legislative history). Though the district court dismissed Plaintiff-Petitioners' claims on this "identicality" ground, other district courts, including those in the Ninth Circuit, have held otherwise, creating what one court has recognized as a "district-court split." *Real World Media LLC v. Daily Caller, Inc.*, __ F. Supp. 3d __, 2024 WL 3835351, at *10 (D.D.C. Aug. 14, 2024).[1] To date, however, no court of appeals has definitively resolved this question. The stakes of this legal

---

[1] The following district court opinions have held, contrary to the district court's conclusion, that § 1202(b) does not impose an identicality requirement: *Real World Media LLC v. Daily Caller Inc.*, __ F. Supp. 3d __, 2024 WL 3835351 (D.D.C. Aug. 14, 2024); *Beijing Meishe Network Tech. Co. v. TikTok Inc.*, No. 23-CV-06012-SI, 2024 WL 3522196 (N.D. Cal. July 23, 2024); *Oracle Int'l Corp. v. Rimini St., Inc.*, No. 2:14-cv-01699-MMD-DJA, 2023 WL 4706127 (D. Nev. July 24, 2023); *Software Pricing Partners, LLC v. Geisman*, No. 3:19-cv-00195-RJC-DCK, 2022 WL 3971292 (W.D.N.C. Aug. 31, 2022); *ADR Int'l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d 411 (S.D. Tex. 2023). The following decisions agree with the district court below: *Andersen v. Stability AI Ltd.*, __ F. Supp. 3d __, 2024 WL 3823234, at *8 (N.D. Cal. Aug. 12, 2024); *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-1931-DMG (EX), 2020 WL 5991503 (C.D. Cal. Aug. 14, 2020); *Frost-Tsuji Architects v. Highway Inn, Inc.*, No. 13-0496 SOM/BMK, 2015 WL 263556 (D. Haw. Jan. 21, 2015).

determination are substantial, both for this case and for others like it, and will determine whether countless copyrights, worth billions of dollars, have any force in the age of generative artificial intelligence ("AI").

Accordingly, the district court here recognized the importance of this question and certified an interlocutory appeal under 28 U.S.C. § 1292(b). This Court should accept jurisdiction and provide a uniform rule—i.e., that there is no identicality requirement § 1202(b)—that will permit district courts to timely and correctly adjudicate the many emerging cases involving the intersection of copyright and generative AI technologies.

II.   RELIEF SOUGHT

Plaintiff-Petitioners respectfully request that this Court grant permission to appeal under 28 U.S.C. § 1292(b) the district court's June 24, 2024 Order, which dismissed Plaintiff-Petitioners' claims under 17 U.S.C. §§ 1202(b)(1) and (b)(3) of the DMCA based on its imposition of an "identicality" requirement. *See* ECF No. 253.

III.   QUESTION PRESENTED

1.   Is liability under § 1202(b) of the DMCA restricted solely to the removal or alteration of Copyright Management Information from an identical copy of a work?

IV. STATEMENT OF FACTS

Plaintiff-Petitioners are software developers. Defendant-Respondent GitHub, owned by Microsoft, supports open-source development of software by creating a repository of code that can be accessed and used by developers. Plaintiff-Petitioners made available on GitHub materials that are subject to various licenses containing conditions for use of those works, most commonly that use of the licensed work requires some form of attribution and copyright notice.

On November 21, 2022, Plaintiff-Petitioners filed their initial complaint, alleging that Defendant-Respondents' violated §§ 1202(b)(1) and (b)(3) by removing or altering CMI from Plaintiff-Petitioners' licensed software code and distributing copies of that code without the requisite CMI. *See* ECF No. 1, ¶¶ 46-77, 138-64. Plaintiff-Petitioners allege that Defendant-Respondents' AI programs, which was trained on code including Plaintiff-Petitioners' which was published with a variety of CMI, emitted copies of the code they were trained on with CMI removed or altered. *See* ECF No. 1, ¶¶ 138–167; *see also* ECF No. 201, ¶¶ 211-35. On May 11, 2023, the District Court permitted Plaintiff-Petitioners' DMCA claims under §§ 1202(b)(1) and (b)(3) to proceed in its order granting in part and denying in part Defendant-Respondents' first motion to dismiss. *See* ECF No. 95, at 18-21.

After Plaintiff-Petitioners filed an amended complaint (ECF No. 97), Defendant-Respondents again moved to dismiss and asked the district court to

reconsider its prior ruling on Plaintiff-Petitioners' § 1202(b) claims, arguing that Defendant-Respondents' copies were not "identical." *See* ECF Nos. 107-2, 109-3.

On January 3, 2024, the district court granted dismissal of Petitioners' section 1202(b) claims, holding that "identical copy[ing]" is required and Petitioners had not sufficiently alleged identical copying. *See* ECF No. 189, at 15. Although the court acknowledged that § 1202(b)(1) prohibits intentional removal or alteration of CMI without the copyright owner's authority, distinct from § 1202(b)(3), which addresses knowing distribution of works with removed or altered CMI, it nonetheless imposed an "identicality" requirement on both claims. *See id.* at 15. It also granted Plaintiff-Petitioners leave to amend. *See id.* at 16-17.

On January 25, 2024, Plaintiff-Petitioners filed their Second Amended Complaint, which included new allegations showing the likelihood that Plaintiff-Petitioners' and class members' licensed code would be emitted verbatim over time as Defendant-Respondents' AI programs' capacity expanded. ECF No. 200. On June 24, 2024, the District Court again dismissed Plaintiff-Petitioners' § 1202(b) claims, ruling that Plaintiff-Petitioners had "failed to meet the DMCA's identicality requirement." ECF No. 253, at 5 (the "MTD Order"). In so doing, it declined to follow *ADR Int'l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d 411 (S.D. Tex. 2023), a recent decision Plaintiff-Petitioners highlighted in their briefing, which holds that the DMCA's prohibition on the removal or alteration of

CMI under § 1202(b) is *not* limited to removal or alteration of CMI from identical copies. *See id* at 425-26.

On July 5, 2024, Plaintiff-Petitioners moved the district court to certify its MTD Order for interlocutory appeal under 28 U.S.C. § 1292(b). ECF No. 268. On September 2, 2024, the district court so certified, concluding that the question of whether section 1202(b) of the DMCA imposes an identicality requirement is a controlling issue of law. *See* ECF No. 282, at 2. The district court found substantial grounds for a difference of opinion on this issue, based on conflicting rulings across various jurisdictions, including decisions in the Ninth Circuit post-dating Plaintiff-Petitioners' motion to certify. *See id.* at 2-3 ("One of the best indications that there are substantial grounds for disagreement on a question of law is that other courts have, in fact, disagreed." (citation omitted)); *see also id.* (collecting cases). Furthermore, the district court also determined that an interlocutory appeal is "likely to materially advance the ultimate outcome of the litigation." *See id.* at 3. The court noted that resolving this legal question at an early stage would prevent the parties from incurring significant litigation costs on issues that might ultimately be rendered moot by a later ruling from the Ninth Circuit. *See id.* at 3. By addressing the controlling legal issue now, the court below concluded that a decision by this Court could avoid unnecessary discovery, expert testimony, and class certification proceedings, ensuring a more efficient path to the

ultimate resolution of the case. *See id.* Accordingly, the district court certified the

issue for appeal and granted a stay of proceedings pending this Court's decision.

See id. at 4.

V.   ARGUMENT

As the district court's certification explains, its order presents controlling

legal questions that are subject to substantial disagreement and will significantly

influence the final resolution of this case. This Court should exercise its discretion

to hear the appeal and resolve this important legal question.

A.   Whether §§ 1202(b)(1) and (b)(3) of the DMCA Impose an "Identicality" Requirement is a Controlling Question of Law

"[A]ll that must be shown in order for a question to be 'controlling' is that

resolution of the issue on appeal could materially affect the outcome of litigation in

the district court." *In re Cement Antitrust Litig. (MDL No. 296)*, 673 F.2d 1020,

1026 (9th Cir. 1982). A "controlling question of law" in an interlocutory appeal is

generally a pure legal question that can be resolved without examining the

particular facts of the case. *Steering Comm. v. United States*, 6 F.3d 572, 575-76

(9th Cir. 1993). In the absence of controlling Ninth Circuit precedent (and indeed,

any appellate court precedent), and in the presence of conflicting district court

authority elsewhere, this Court should review pure questions of law such as

whether § 1202(b) requires CMI be removed or altered from identical copies. *See*

id.

Whether § 1202(b) requires parties to plead and prove "identicality" between the original and the copy from which CMI was removed or altered is a pure legal question of statutory interpretation. Such issues are particularly appropriate for interlocutory appeal. *See Joffee v. Google, Inc.*, 746 F.3d 920, 923-24 (9th Cir. 2013) (interlocutory appeal certified under 28 U.S.C. § 1292(b) because the district court "resolved a novel question of statutory interpretation"); *see also Sec'y, U.S. Dep't of Lab. v. Preston*, 873 F.3d 877, 880 (11th Cir. 2017) (conducting *de novo* review because "[t]his interlocutory appeal presents a purely legal issue of statutory interpretation"); *Fast v. Applebee's Int'l, Inc.*, 638 F.3d 872, 785-76 (8th Cir. 2011) (conducting *de novo* review of district court's statutory interpretation on interlocutory appeal).

B.   Existing Conflicts Over § 1202(b)'s "Identicality" Requirement Highlight the Urgent Need for Appellate Resolution

Section 1292(b)'s requirement for a "substantial ground for difference of opinion" is easily satisfied here. Courts generally find a substantial ground for difference of opinion exists when "other courts have, in fact, disagreed." *Rollins v. Dignity Health,* No. 13-CV-01450-TEH, 2014 WL 6693891, at *3 (N.D. Cal. Nov. 26, 2014) (finding "substantial grounds for disagreement" where "two district courts have decided this issue explicitly in conflict with this Court's decision"); *see also Couch v. Telescope Inc.*, 611 F.3d 629, 633 (9th Cir. 2010) (quoting 3 Federal Procedure, Lawyers Edition § 3:212 (2010) ("the circuits are in dispute on the

question and the court of appeals of the circuit has not spoken on the point, if complicated questions arise under foreign law, or if novel and difficult questions of first impression are presented.")). This Court has confirmed that "when novel legal issues are presented, on which fair-minded jurists might reach contradictory conclusions, a novel issue may be certified for interlocutory appeal without first awaiting development of contradictory precedent." *Reese v. BP Expl. (Alaska) Inc.*, 643 F.3d 681, 688 (9th Cir. 2011). The identification of conflicting or contradictory rulings provides sufficient grounds for a substantial difference of opinion. *See Couch*, 611 F.3d at 633-34 (citing *Union Cnty., Iowa v. Piper Jaffray & Co., Inc.*, 525 F.3d 643 (8th Cir. 2008)).

Whether §§ 1202(b)(1) or (b)(3) requires that CMI be removed or altered from an identical copy of a work is a novel legal issue, one on which no Court of Appeals has yet directly spoken. District courts in this Circuit and across the country likewise are sharply divided on this question.

The court below primarily relied on district court opinions such as *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-1931-DMG (EX), 2020 WL 5991503, at *6 (C.D. Cal. Aug. 14, 2020) which have held that "identicality" is required for claims under §§ 1202(b)(1) and (b)(3) to proceed. *See also Kelly v. Arriba Soft Corp.,* 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999); *Frost-Tsuji Architects v. Highway Inn, Inc.*, No. 13-00496 SOM/BMK, 2015 WL 263556, at

*2 (D. Haw. Jan. 21, 2015); *Fischer v. Forrest*, 286 F. Supp. 3d 590 (S.D.N.Y. 2018). On the other hand, district courts, including those within the Ninth Circuit, have held that DMCA liability can attach even when the work in question is not an "exact copy" or has been "altered," thereby implying that the copied work need not be identical. *See, e.g.*, *ADR Int'l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d 411 (S.D. Tex. 2023). For example, in *Oracle Int'l Corp. v. Rimini St., Inc.*, No. 2:14-cv-01699-MMD-DJA, 2023 WL 4706127, at *82 (D. Nev. July 24, 2023), the court expressly rejected the argument "that a work that removes copyright management information must be an exact copy of the original work." Similarly, in *Software Pricing Partners, LLC v. Geisman*, No. 3:19-cv-00195-RJC-DCK, 2022 WL 3971292, at *5 (W.D.N.C. Aug. 31, 2022), the court found DMCA liability where the defendant "altered SPP's copyrighted documents by removing indications of the copyright or otherwise altering the documents prior to distributing [them] to customers," despite those alterations. The court emphasized that Geisman, a former employee, knowingly altered the documents to make them appear as his own while maintaining a substantial similarity to the original work.

The district court's adoption of an "identicality" requirement for § 1202(b) claims in this case is also in conflict with other cases addressing DMCA violations in the context of software code. In such cases, courts have recognized that CMI can be embedded within the code itself, and its removal would inherently result in an

9

infringing copy that is no longer identical to the original. In *Oracle Int'l*, 2023 WL 4706127, at *82, the court held that when a defendant modifies source code that is "substantially similar" to the plaintiff's copyrighted code, including by replacing the author's name with its own, the defendant can be held liable under the DMCA. *See also Bounce Exch., Inc. v. Zeus Enter. Ltd.*, No. 15cv3268 (DLC), 2015 WL 8579023, at *3 (S.D.N.Y. Dec. 9, 2015) (finding § 1202(b) liability for removing CMI woven into and incorporated within the code). Moreover, at least one court in this circuit has explicitly recognized that a § 1202(b) claim is viable when the infringing copy is a "derivative"—which, by definition, is not identical. *See Splunk, Inc. v. Cribl, Inc.*, 662 F. Supp. 3d 1029, 1053-54 (N.D. Cal. 2023).

Within a day of Plaintiff-Petitioners' filing their motion seeking certification for interlocutory appeal below, another court in the Northern District of California denied a motion to dismiss a § 1202(b) claim and rejected that an element of identicality is required to state a § 1202(b) claim. *Beijing Meishe Network Tech. Co. v. TikTok Inc.*, No. 23-CV-06012-SI, 2024 WL 3522196, at *9 (N.D. Cal. July 23, 2024). *Meishe* acknowledged a "split in authority" as to § 1202(b)'s alleged "identicality" standard, and countered the district court's adoption of an "identicality" standard in this action with two recent Ninth Circuit district court cases holding the opposite. *See id.* at *9 (comparing the district court's order with *Oracle Int'l Corp.*, 2023 WL 4706127, at *82, and *Splunk Inc.*, 662 F. Supp. 3d at

1054). Even though the *Meishe* plaintiffs, much like Plaintiff-Petitioners here,

alleged that the copied source code was only "strikingly similar" to the original

works because the copies comprised portions of the original works and contained

minor typographical variations and obviously altered CMI, the *Meishe* court held

that plaintiffs had sufficiently pled a § 1202(b) claim. *See id.*

Three weeks later, on August 12, another court in the Northern District of

California granted a motion to dismiss a DMCA claim involving generative AI

technology on the ground that plaintiffs' allegations failed to satisfy § 1202(b)'s

purported "identicality" standard. *See Andersen*, 2024 WL 3823234, at *8. While

the *Andersen* court "agreed with the reasoning" of the district court in this action

regarding the "identicality" standard, he also recognized the very same "split in

authority" identified by the *Meishe* court, describing § 1202(b)'s "identicality"

standard as an "issue [that] is unsettled." *Id.*

Just two days after *Andersen*, a court in the District Court of the District of

Columbia, while recognizing a "nascent district-court split" regarding 1202(b)'s

"identicality" requirement rejected the argument that "exact" copies are required to

support a DMCA claim. *Real World Media*, 2024 WL 3835351, at *10 (collecting

cases). Notably, the *Real World Media* court held that "nothing in § 1202(b)

requires precise equivalence between the work from which CMI is removed and

the allegedly infringing work," and that a § 1202(b) action can lie even where a

party copies portions of a work rather than the entire work. *Id.*

The foregoing demonstrates that reasonable jurists could reach—and indeed have reached—different conclusions regarding the issue of "identicality" for § 1202 liability. The difficulty of the issue, the absence of guidance from any circuit court—particularly the Ninth Circuit—and the conflicting rulings among district courts all underscore that "substantial grounds for difference of opinion" exist. *See, e.g.*, *Kinghoffer v. S.N.C. Achille Lauro Ed Altri-Gestone Motonave Achille Lauro In Amministrazione Straordinaria, et al.*, 921 F.2d 21, 25 (2d Cir. 1990); *Shoreham Co-op. Apple Producers Ass'n, Inc. v. Donovan*, 764 F.2d 135, 140 n.14 (2d Cir. 1985) (noting that "there were substantial grounds for difference of opinion" because of a "difference of opinion among the district courts").

C. Immediate Appellate Review Will Streamline This Litigation and Provide Crucial Guidance for Similar Cases in This Circuit and Elsewhere

As this Court has observed, "[n]either § 1292(b)'s literal text nor controlling precedent requires that the interlocutory appeal have a final, dispositive effect on the litigation, only that it 'may materially advance' the litigation." *Reese*, 643 F.3d at 688 (citing 28 U.S.C. § 1292(b)). Settling the legal standard at issue here not only would materially advance this class action, but also would impact numerous other challenges to other AI models across the country. *J. B. v. G6 Hosp., LLC,* No. 19-CV-07848-HSG, 2021 WL 6621068, at *4 (N.D. Cal. Dec. 16, 2021) ("Rather

than litigating the case to the finish under a standard that will be challenged on appeal, the Court and the parties will benefit from definitive guidance from the Ninth Circuit at the outset, before time and resources are invested."). Certification will materially advance this litigation for at least four reasons:

First, courts have consistently certified questions for interlocutory appeal when the issues go to the heart of the case. *See, e.g.*, *Sterk v. Redbox Automated Retail, LLC*, 672 F.3d 535, 536 (7th Cir. 2012) (recognizing the importance of resolving claims that represent the "main" issues in the case). Plaintiff-Petitioners' § 1202(b) claims—their remaining federal claims—are undeniably central to this litigation. *Cf. U.S. ex rel Huangyan Import & Export Corp., v. Nature's Farm Prod., Inc.*, 370 F. Supp. 2d 993, 1005 (N.D. Cal. 2005) ("Depending on how the three issues are resolved [on interlocutory appeal], the United States might have two FCA claims, one FCA claim, no claim at all or might be in the wrong court altogether."). The outcome of this appeal will dictate the fate of those claims. The monetary stakes are equally substantial: if Plaintiff-Petitioners succeed on their DMCA claim, they are entitled to statutory damages between $2,500 and $25,000 per violation. 17 U.S.C. § 1206(c)(3)(B). With Plaintiff-Petitioners alleging tens of thousands, if not more, individual violations, potential damages could reach billions of dollars.

Second, and closely related, this is a class action. Granting an interlocutory

appeal would "shorten the period between the commencement of the action and its ultimate termination and would avert unnecessary work and expense," an obvious efficiency gain, particularly in class actions, where litigants would be better positioned to predict their likelihood of success, the scope of potential liability, and a fair estimate of the case's value. *See In re Cement Antitrust Litig.*, 673 F.2d at 1029 (discussing legislative history of 1292(b)) (Boochever, J., dissenting); 9 Moore, Federal Practice, P 110.22(2), at 260 (2d ed. 1975) ("The critical requirement is that it (the question) have the potential for substantially accelerating the disposition of the litigation."). As observed by another court in this circuit, "especially in class actions, uncertainty over a key claim's status may delay settlement (almost all class actions are settled . . .), and by doing so further protract the [case]." *Casa v. Victoria's Secret Stores, LLC*, No. 14-6412-GW(VBKx), 2015 WL 13446989, at *3 (C.D. Cal. April 4, 2015) (quoting *Sterk*, 672 F.3d at 536 (Posner, J.)). "That is enough to satisfy the 'may materially advance' cause of section 1292(b)." *Sterk*, 672 F.3d at 536.

Similarly, in *Canela v. Costco Wholesale Corp.*, No. 13-cv-03598-BLF, 2018 WL 3008532, at *2 (N.D. Cal. June 15, 2018), the court recognized that resolving a key legal issue can lead to a more efficient trial or settlement. *Id.* ("If *Canela* is limited to pursuing only her individual PAGA claim . . . the trial would involve fewer disputed issues and it would be more likely that the parties would

reach a settlement."). Likewise, *Rollins*, 2014 WL 6693891, at *4, affirmed that

addressing critical legal questions early "saves time and expense" and may

"encourage a negotiated settlement," which could entirely resolve the litigation.

Third, courts have consistently found this requirement satisfied when the

resolution of an interlocutory appeal would impact a significant number of other

cases. *See, e.g.*, *Klinghoffer v. S.N.C. Achille Lauro Ed Altri-Gestione Motonave

Achille Lauro In Amministrazione Straordinaria*, 921 F.2d 21, 24 (2d Cir. 1990)

("[T]he impact that an appeal will have on other cases is a factor that we may take

into account in deciding whether to accept an appeal that has been properly

certified by the district court."); *see also Leite v. Crane Co.*, No. 11–00636

JMS/RLP, 2012 WL 1982535, at *7 (D. Haw. May 31, 2012) (collecting cases).

Given the number of other cases challenging AI models—many of them class

actions—asserting similar DMCA claims, a resolution of these issues on

interlocutory appeal will benefit those similarly situated. *See Ass'n of Irritated

Residents v. Fred Schakel Dairy*, 634 F. Supp. 2d 1081, 1093 (E.D. Cal. 2008)

("The opportunity to achieve appellate resolution of an issue important to other

similarly situated [parties] can provide an additional reason for certification."

(citing, *inter alia*, *Klinghoffer*, 921 F.2d at 24)).

Moreover, the Ninth Circuit's intervention is necessary to prevent

conflicting interpretations of the DMCA across federal courts. As AI-related cases

continue to proliferate, the risk of courts reaching differing conclusions will grow. Without guidance, both defendants and plaintiffs will face a patchwork of standards, leading to inconsistent outcomes and unnecessary legal complexity in an area critical to the future of digital copyright enforcement.

Fourth, this case remains at an early procedural stage. Given the significance of Plaintiff-Petitioners' § 1202(b) claims, resolving the legal question now would avoid duplicative efforts should the district court's dismissal of these claims be reversed after trial. *See Casas*, 2015 WL 13446989, at *4. At this point, discovery has not materially progressed, meaning that little effort would need to be duplicated, and in the event of reversal, much future duplicative work could be avoided. *See Finder v. Leprino Foods Co.,* No. 1:13-CV-02059 AWI BAM, 2016 WL 4095833, at *4 (E.D. Cal. Aug. 1, 2016). Conversely, if this case proceeds and the dismissal is later reversed, both the parties and the court below will face substantial duplication of effort. *See Sterk*, 672 F.3d at 536.

D. The District Court's Rewriting of § 1202(b) to Require "Identicality" Was Error

Interlocutory review is also warranted because the district court erred in resolving the controlling question of law at issue.

The district court's holding contravenes the plain text of the statute. *See Coronado-Durazo v. I.N.S.*, 123 F.3d 1322, 1325 (9th Cir. 1997) ("[U]nder the established approach to statutory interpretation, we rely on plain language in the

first instance[.]"). It is a standard principle of statutory construction that identical

words and phrases should normally be given the same meaning. *Powerex Corp. v.*

Reliant Energy Servs., Inc., 551 U.S. 224, 232 (2007). The text of § 1202(b) does

not include the word "identical" or anything approximating it. *See Transwestern*

Pipeline Co., LLC v. 17.19 Acres of Prop. Located in Maricopa Cnty., 627 F.3d

1268, 1270 (9th Cir. 2010); *Coronado-Durazo,* 123 F.3d at 1325 ("[U]nder the

established approach to statutory interpretation, we rely on plain language in the

first instance[.]"). On the other hand, the DMCA mentions the word "identical"

only once elsewhere in the statute. The word "identical" is used in a provision

exempting from liability nonprofit libraries, archives, and educational institutions

under § 1201. *See* 17 U.S.C. § 1201(d)(2) ("The exemption made available under

paragraph (1) shall only apply with respect to a work when an *identical* copy of

that work is not reasonably available in another form." (emphasis added)). The

inclusion of "identical" in § 1202's sister provision shows that if Congress had

intended to require "identicality" for § 1202(b) liability, it would have said so.[2]

That Congress omitted the word "identical" in § 1202 is meaningful. *See Keene*

Corp. v. United States, 508 U.S. 200, 208 (1993) ("[W]here Congress includes

---

[2] Though the language of the statute is unambiguous, the legislative history of the
DMCA is in accord. *S.E.C. v. McCarthy*, 322 F.3d 650, 655 (9th Cir. 2003). There
is absolutely nothing in the legislative materials leading to the enactment of the
DMCA that indicates Congress intended to impose an "identicality" requirement
under § 1202(b). *See* S. Rep. No. 105-190, at 31 (1998).

particular language in one section of a statute but omits it in another, it is generally presumed that Congress acts intentionally and purposely in the disparate inclusion or exclusion.") (internal quotation marks omitted, citing *Russello v. United States*, 464 U.S. 16, 23 (1983)); *Pettis ex rel. U.S. v. Morrison-Knudsen Co., Inc.,* 577 F.2d 668, 672 (9th Cir. 1978) ("We have no doubt but that under such circumstances the intent of Congress resides in the words of the statute. That is, discharge of our obligation to follow the intent of Congress requires that we assume that Congress said what it meant and meant what it said.").

Further, to the extent the district court relied on district court cases such as *Kirk Kara*, to impute an "identicality" requirement into § 1202, that too was error. "[A]lthough the court in *Kirk Kara* held the DMCA requires identical copies, the case law it cited does not support its holding." *ADR Int'l*, 667 F. Supp. 3d at 427. For example, *Kirk Kara* referenced *Kelly,* 77 F. Supp. 2d at 1122, and *Frost-Tsuji Architects*, 2015 WL 263556, at *2, but neither of those cases mentioned or imposed an identical-copies requirement under the DMCA. Similarly, *Kirk Kara* cited *Fischer*, 286 F. Supp. 3d 590, where the term "identical" does not appear at all. In fact, *Fischer* did not endorse such a requirement." *Id.* at 609 (finding the plaintiffs failed to plead that the "underlying work ha[d] been substantially or entirely reproduced").

V.   CONCLUSION

For the reasons outlined above, Plaintiff-Petitioners respectfully request that this Court grant permission to appeal under 28 U.S.C. § 1292(b) to allow this Court to definitively address whether §§ 1202(b)(1) and (b)(3) of the DMCA impose an "identicality" requirement for liability.

Dated: October 7, 2024   Respectfully submitted,

/s/ Joseph R. Saveri
Joseph R. Saveri
Cadio Zirpoli
Christopher K. L. Young
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1505
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
Email: czirpoli@saverilawfirm.com
Email: cyoung@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
Telephone: (323) 968-2632
Facsimile: (415) 395-9940
Email: mb@butticklaw.com

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

Jesse Panuccio
BOIES SCHILLER FLEXNER LLP
1401 New York Avenue, NW
Washington, DC 20005
Telephone: (202) 237-2727
Email: jpanuccio@bsfllp.com

*Attorneys for Plaintiff-Petitioners and the
Proposed Class*

CERTIFICATE OF COMPLIANCE

I hereby certify that this brief complies with Fed. R. App. P. 5(c)(1) because it contains no more than 5,200 words, excluding the portions of the brief exempted by Rule 32(f), according to the count of Microsoft Word. I further certify that this brief complies with the typeface and type-style requirements of Rule 32(a)(5)-(6) because it is printed in a proportionally spaced 14-point font, Times New Roman.

Dated: October 7, 2024   Respectfully submitted,

           /s/ Joseph R. Saveri
            Joseph R. Saveri

CERTIFICATE OF SERVICE

I hereby certify that on October 7, 2024, I electronically filed and served the foregoing Petition For Permission To Appeal Order Certified Under 28 U.S.C. § 1292(b) by using this Court's electronic-filing system.

Dated: October 7, 2024   Respectfully submitted,

                                   /s/ Joseph R. Saveri
                                      Joseph R. Saveri

EXHIBIT A

1

2

3

4   UNITED STATES DISTRICT COURT

5   NORTHERN DISTRICT OF CALIFORNIA

6

7   J. DOE 1, et al.,   Case No. 22-cv-06823-JST

8   Plaintiffs,

9   v.   **ORDER GRANTING IN PART
                                                   DENYING IN PART MOTIONS TO
10 GITHUB, INC., et al.,   DISMISS**

                     Defendants.   Re: ECF Nos. 215, 219
11

12

13   Before the Court are motions to dismiss filed by Defendants GitHub, Inc., and Microsoft

14 Corporation (collectively, "Defendant GitHub"), ECF No. 215; and Defendants OpenAI, Inc.,

15 OpenAI, L.P., OpenAI OPCO, L.L.C., OpenAI GP, L.L.C., OpenAI Startup Fund GP I, L.L.C.,

16 OpenAI Startup Fund I, L.P., and OpenAI Startup Fund Management, LLC (collectively,

17 "Defendant OpenAI"), ECF No. 219. The Court will grant the motions in part and deny them in

18 part.[1]

19 **I.   BACKGROUND**

20   Because the facts are well-known to the parties and the Court has summarized Plaintiffs'

21 allegations in detail in its prior orders, *see* ECF Nos. 95, 189, the Court will not elaborate them

22 here.

23   Following the last round of briefing, the Court found that Plaintiffs alleged standing for

24 damages as to Does 1, 2, and 5, but not Does 3 and 4. Further, the Court dismissed Plaintiffs'

25 state law claims for intentional and negligent interference with prospective economic relations,

26

27 ───────────────

28 [1] Although the caption of this order cites to the redacted version of Defendant GitHub's motion to
    dismiss, ECF No. 215, the remainder of the order will refer to the sealed version of this
    document—ECF No. 214-2.

United States District Court
Northern District of California

1   unjust enrichment, negligence, and unfair competition with prejudice. And finally, the Court

2   dismissed Plaintiffs' claim under Section 1202(b) of the Digital Millenium Copyright Act

3   ("DMCA"), reasoning that Plaintiffs failed to meet Section 1202(b)'s identicality requirement.

4   "[O]ut of abundance of caution," however, Plaintiffs were granted leave to amend their DMCA

5   claim. ECF No. 189 at 16.

6   In Plaintiffs' second amended complaint ("SAC"), ECF No. 201, three claims remain.[2]

7   Count One alleges a violation of DMCA Section 1202(b)(1) and 1202(b)(3) against all

8   Defendants. ECF No. 201 at 53. Count Two alleges breach of contract for violation of open-

9   source licenses against all Defendants. *Id.* at 59. Count Three alleges breach of contract for

10   selling licensed materials against only Defendant GitHub. *Id.* at 61.

11   In support of these claims, Plaintiffs add two primary new assertions to their SAC. First,

12   they allege that "[i]n July 2022," "GitHub introduced a user-settable Copilot filter called

13   '[s]uggestions matching public code.'" *Id.* ¶ 145. This filter is also referred to as the

14   "duplication-detection feature." *Id.* ¶ 146 n.23. Users can set the filter "to either allow or block

15   code completion suggestions that match publicly available code." *Id.* ¶ 146. If a user chooses to

16   block suggestions that match public code, "GitHub Copilot checks code completion suggestions

17   with their surrounding code of about 150 characters against public code on GitHub." *Id.* "If there

18   is a match, or a near match, the suggestion is not shown" to the user. *Id.* However, "GitHub

19   makes [this feature] entirely optional to users, and provides no such optionality to licensors." *Id.* ¶

20   148. Therefore, in Plaintiffs' telling, "users who want to receive identical code from GitHub or do

21   not want to exclude it, may do so." *Id.* In light of this feature, Plaintiffs "believe it is likely that

22   their licensed code is omitted[3] by Github [sic] in violation of the open source licenses[,]" and that

23   "there is a substantial risk, if not certainty, that identical code will be emitted in the future." *Id.* ¶

24   149.

25   Second, Plaintiffs allege that "[r]ecent academic research shows that the likelihood

26   Plaintiffs' or class members' code would be emitted verbatim is only increasing." *Id.* ¶ 104. They

27

28   [2] The redacted version of Plaintiffs' SAC is available at ECF No. 200.
       [3] For purposes of this order, the Court assumes Plaintiffs meant "emitted."

1   cite to a study entitled *Quantifying Memorization Across Neural Language Models* by Nicholas

2   Carlini et al., which reasoned that "[m]emorization significantly grows as we increase (1) the

3   capacity of a model, (2) the number of times an example has been duplicated, and (3) the number

4   of tokens of context used to prompt the model."[4] *Id.* Accordingly, Plaintiffs assert that "as

5   generative AI models such as Copilot increase capacity and continue to scale, it becomes more

6   likely that training data will become memorized and emitted verbatim, i.e., as an exact duplicate."

7   *Id.* ¶ 105.

8   Both Defendants move to dismiss Plaintiffs' Section 1202(b) claim pursuant to Fed. R.

9   Civ. P. 12(b)(6). In addition, OpenAI moves to dismiss Plaintiffs' breach of contract claim for

10 violation of open-source licenses, and GitHub moves to dismiss Plaintiffs' requests for unjust

11 enrichment and punitive damages.

II.   JURISDICTION

13   The Court has jurisdiction over Plaintiffs' federal claims pursuant to 28 U.S.C. § 1331 and

14 supplemental jurisdiction over Plaintiffs' state law claims under 28 U.S.C. § 1367.

III.   LEGAL STANDARD

16   **A.   Rule 12(b)(6)**

17   "Dismissal under Rule 12(b)(6) is appropriate only where the complaint lacks a cognizable

18 legal theory or sufficient facts to support a cognizable legal theory." *Mendiondo v. Centinela*

19 *Hosp. Med. Ctr.*, 521 F.3d 1097, 1104 (9th Cir. 2008). To survive a motion to dismiss, "a

20 complaint must contain sufficient factual matter, accepted as true, to 'state a claim to relief that is

21 plausible on its face.'" *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) (quoting *Bell Atlantic Corp. v.*

22 *Twombly*, 550 U.S. 544, 570 (2007)). "A claim has facial plausibility when the plaintiff pleads

23 factual content that allows the court to draw the reasonable inference that the defendant is liable

24 for the misconduct alleged." *Id*. In determining whether a plaintiff has met the plausibility

25 requirement, a court must "accept all factual allegations in the complaint as true and construe the

26

27 [4] The full citation of this study is as follows: Nicholas Carlini, et al., *Quantifying Memorization Across Neural Language Models*, arXiv (submitted Feb. 15, 2022, revised Mar. 6, 2023),

28 https://arxiv.org/pdf/2202.07646.pdf (last accessed April 16, 2023). Going forward, the Court refers to this study as the "Carlini Study."

United States District Court
Northern District of California

1   pleadings in the light most favorable to the nonmoving party." *Knievel v. ESPN*, 393 F.3d 1068,

2   1072 (9th Cir. 2005).

3   **B.**   **Leave to Amend**

4   Leave to amend a complaint "shall be freely given when justice so requires." Fed. R. Civ.

5   P. 15(a)(2). The decision of whether to grant leave to amend is "within the discretion of the

6   district court, which may deny leave due to 'undue delay, bad faith or dilatory motive on the part

7   of the movant, repeated failure to cure deficiencies by amendments previously allowed, undue

8   prejudice to the opposing party by virtue of allowance of the amendment, and futility of

9   amendment.'" *Leadsinger, Inc. v. BMG Music Pub.*, 512 F.3d 522, 532 (9th Cir. 2008) (quoting

10 *Foman v. Davis*, 371 U.S. 178, 182 (1962)).

11 **IV.**   **DISCUSSION**

12   **A.**   **DMCA Section 1202(b)(1) and 1202(b)(3)**

13   Defendants ask the Court to dismiss Plaintiffs' Section 1202(b) claim. *See* ECF Nos. 214-

14 2 at 17–22, 219 at 10–13. Although each Defendant contends that Plaintiffs' claim fails on

15 various grounds, the Court finds one argument dispositive: Plaintiffs again fail Section 1202(b)'s

16 identicality requirement.

17   Plaintiffs' opposition spills much ink arguing that identicality is not an element of a

18 Section 1202(b) claim. *See* ECF Nos. 234 at 12–16, 235 at 12–15. Having twice addressed this

19 issue already, the Court will not revisit it at length.[5] Plaintiffs focus on a non-binding decision

20 from the Southern District of Texas, *ADR Int'l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d

21 411, 425 (S.D. Tex. 2023), which concluded that the "DMCA is not limited to [copyright

22 management information] conveyed in connection with identical copies of a work." But caselaw

23 from courts in the Ninth Circuit continues to compel this Court to reach a different conclusion.

24 *See, e.g.*, *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-1931, 2020 WL 5991503, at *6

25 (C.D. Cal. Aug. 14, 2020) ("courts have found that no DMCA violation exists where the works are

26

27 [5] On February 28, 2024, Plaintiffs filed a motion for leave to file a motion for reconsideration of
the Court's prior order, ECF No. 189, pursuant to Civil Local Rule 7–9. ECF No. 218. The Court

28 ordered Defendants to file responses by March 15, 2024. After considering arguments from both
sides, the Court declined to grant Plaintiffs' motion. ECF No. 246.

United States District Court
Northern District of California

1   not identical."); *Advanta-STAR Auto. Rsch. Corp. of Am. v. Search Optics, LLC*, 672 F. Supp. 3d

2   1035, 1057 (S.D. Cal. 2023) ("Plaintiff has not plausibly alleged that Defendants distributed

3   identical copies of Plaintiff's comparison"); *Frost-Tsuji Architects v. Highway Inn, Inc.*, No. CIV.

4   13-00496 SOM, 2015 WL 263556, at *3 (D. Haw. Jan. 21, 2015), *aff'd*, 700 F. App'x 674 (9th

5   Cir. 2017) (denying Section 1202(b) claim where the drawing at issue was "not identical" to the

6   drawing by plaintiff); *Tremblay v. OpenAI, Inc.*, No. 23-CV-03223-AMO, — F.Supp.3d —, 2024

7   WL 557720, at *5 (N.D. Cal. Feb. 12, 2024) (citing *Kirk Kara Corp.* for the proposition that a

8   defendant must make identical copies of a plaintiff's work "to implicate the DMCA.").

9   Turning to the allegations in the SAC, Defendants argue that "[l]ike the FAC, the SAC

10   does not identify even a single example of Copilot producing an identical copy of any work."

11   ECF No. 219 at 10. The Court agrees. The SAC includes the same allegations that Defendants'

12   programs released, or "output," code published to GitHub by Does 1, 2, and 5. *See* ECF No. 201

13   ¶¶ 115, 120, 121, 124, 125, 133. In its last order, the Court concluded that these facts were "not

14   sufficient for a Section 1202(b) claim" because they were not identical. ECF No. 189 at 15.

15   Because these facts have not changed, the Court must again conclude that Plaintiffs have failed to

16   meet the DMCA's identicality requirement.

17   Plaintiffs' new allegations fare no better. Plaintiffs allege that, should a user elect to *not*

18   use the duplication-detection feature, a user could conceivably view an identical match of 150

19   characters, and use it without attribution. *See* ECF No. 201 ¶¶ 147–151. While Plaintiffs aver that

20   the duplication-detection tool "by definition establishes Copilot's ability to reproduce verbatim

21   copies of code," ECF No. 234 at 18 (emphasis omitted), they "do not explain how the tool makes

22   it plausible that Copilot will in fact do so through its normal operation or how any such verbatim

23   outputs are likely to be anything beyond short and common boilerplate functions." ECF No. 242-

24   1 at 16; *see also Iqbal*, 556 U.S. at 680 (finding that plaintiff had not "nudged [his] claims" . . .

25   "across the line from conceivable to plausible."). And, as Defendant GitHub also points out, "the

26   mere existence of such a feature does not make it more likely that Copilot would ever output an

27   identical copy of *Plaintiffs'* works." ECF No. 214-2 at 21 (emphasis in original). Thus, Plaintiffs

28   have not demonstrated that Copilot's duplication-detection tool is likely to give rise to Section

1   1202(b) liability.

2   In addition, the Court is unpersuaded by Plaintiffs' reliance on the Carlini Study. It bears

3   emphasis that the Carlini Study is not exclusively focused on Codex or Copilot, and it does not

4   concern Plaintiffs' works. That alone limits its applicability. And further, as Defendant GitHub

5   notes, the Carlini Study does nothing to "rehabilitate Plaintiffs' own concession that, still, 'more

6   often,' Copilot's suggestions are 'a modification.'" ECF No. 214-2 at 21 (quoting ECF No. 201 ¶

7   108).

8   The Study "tested multiple models by feeding prefixes of prompts based on training data

9   into each model in order to compare the performance of models of different sizes to emit output

10   that is identical to training data." ECF No. 201 ¶ 104. It determined that when models are

11   "prompted appropriately, they will emit the memorized training data verbatim." *Id.* (quoting

12   Carlini Study). In regard to the GitHub Copilot model in particular, the Study concluded that it

13   "rarely emits memorized code in benign situations, and most memorization occurs only when the

14   model has been prompted with long code excerpts that are very similar to the training data."

15   Carlini Study at 6. To paraphrase Defendant GitHub, "Plaintiffs tried to [prompt Copilot] in their

16   last complaint . . . to generate an identical copy of their code" and they were unable to do so. ECF

17   No. 214-2 at 22 (emphasis omitted). Accordingly, Plaintiffs' reliance on a Study that, at most,

18   holds that Copilot may theoretically be prompted by a user to generate a match to someone else's

19   code is unpersuasive.

20   To conclude, the Court dismisses Plaintiffs' Section 1202(b) claim. Having previously

21   dismissed this claim on the same ground, the Court will now dismiss Plaintiffs' Section 1202(b)

22   claim with prejudice.

23   **B.   Breach of Contract**

24   Defendant OpenAI moves to dismiss Plaintiffs' breach of contract claim for violation of

25   open-source licenses. ECF No. 219 at 13. In support of this argument, OpenAI contends that (1)

26   Plaintiffs fail to state a claim based on Codex; (2) Plaintiffs fail to state a claim based on Copilot;

27   and (3) Plaintiffs' theory based on Copilot fails on the merits because "the attribution and notice

28   terms are conditions and do not give rise to a contract claim." *Id.* at 14–17. Plaintiffs respond that

6

1   OpenAI has waived its right to challenge Plaintiffs' breach of contract claim pursuant to Federal

2   Rule of Civil Procedure 12(g)(2), and in any event, its arguments fail on the merits. Having

3   considered these arguments, the Court declines to dismiss Plaintiffs' breach of contract claim.

4   **1.   Rule 12(g)(2)**

5   The Court begins by examining whether Rule 12(g)(2) forecloses OpenAI's arguments.

6   Rule 12(g)(2) states: "Except as provided in Rule 12(h)(2) or (3), a party that makes a motion

7   under this rule must not make another motion under this rule raising a defense or objection that

8   was available to the party but omitted from its earlier motion." Fed. R. Civ. P. 12(g)(2). The

9   Ninth Circuit has made clear that it "read[s] Rule 12(g)(2) in light of the general policy of the

10   Federal Rules of Civil Procedure, expressed in Rule 1. That rule directs that the Federal Rules 'be

11   construed, administered, and employed by the court and the parties to secure the just, speedy, and

12   inexpensive determination of every action and proceeding.'" *In re Apple iPhone Antitrust Litig*.,

13   846 F.3d 313, 318 (9th Cir. 2017), *aff'd sub nom. Apple Inc. v. Pepper*, 139 S. Ct. 1514 (2019)

14   (quoting Fed. R. Civ. P. 1). Accordingly, "[d]enying late-filed Rule 12(b)(6) motions . . . can

15   produce unnecessary and costly delays, contrary to the direction of Rule 1." *Id.*; *see also Banko v.*

16   *Apple, Inc*., No. 13–02977 RS, 2013 WL 6623913, at *2 (N.D. Cal. Dec. 16, 2013) (internal

17   quotations omitted) ("Although Rule 12(g) technically prohibits successive motions to dismiss that

18   raise arguments that could have been made in a prior motion . . . courts faced with a successive

19   motion often exercise their discretion to consider the new arguments in the interests of judicial

20   economy."). The Court agrees with OpenAI that "[c]onsidering [these] arguments now could

21   'materially expedite[] the district court's disposition of the case,' and avoid unnecessarily delay

22   and cost to both parties related to the contract claim, consistent with the direction of Rule 1." ECF

23   No. 244 at 15 (quoting *In re Apple*, 846 F.3d at 320). It will thus consider OpenAI's arguments.

24   **2.   Plaintiffs State a Claim Based on Codex**

25   OpenAI's first argument is that Plaintiffs fail to state a claim based on Codex. OpenAI

26   contends that "[t]he SAC does not identify the contracts at issue or explain how the Doe Plaintiffs

27   and OpenAI entered those contracts, how OpenAI purportedly breached those contracts with

28   respect to the Doe Plaintiffs, or how Doe Plaintiffs suffered damage from that breach." ECF No.

United States District Court
Northern District of California

7

1   244 at 15 (citing *Gautier v. Gen. Tel. Co.*, 234 Cal. App. 2d 302, 305–06 (Ct. App. 1965)).

2   In its prior order denying Defendants' motions to dismiss Plaintiffs' breach of contract

3   claim, the Court explained:

> Plaintiffs advance claims for breach of the eleven suggested licenses
> GitHub presents to users that require (1) attribution to the owner, (2)
> inclusion of a copyright notice, and (3) inclusion of the license
> terms. [ECF No. 1] ¶ 34 n.4. [. . .] Plaintiffs allege that use of
> licensed code "is allowed only pursuant to the terms of the
> applicable Suggested License," and that each such license requires
> that any derivative work or copy include attribution, a copyright
> notice, and the license terms. *Id*. ¶¶ 173, 34 n.4. Plaintiffs further
> allege that Codex and Copilot reproduce licensed code as output
> without attribution, copyright notice, or license terms, thereby
> violating the relevant provisions of each license. While Plaintiffs do
> not identify the specific subsections of each suggested license that
> correspond to each of these requirements, the Court finds that
> Plaintiffs have sufficiently identified "the contractual obligations
> allegedly breached," as required to plead a breach of contract claim.
> [*Williams v. Apple, Inc*., 449 F. Supp. 3d 892, 908 (N.D. Cal.
> 2020)].

13   ECF No. 95 at 22. Having already determined that Plaintiffs adequately stated a breach of

14   contract claim, the Court declines to reanalyze this issue. Therefore, Plaintiffs' breach of contract

15   claim will not be dismissed on this ground.

16   **3.   Plaintiffs State a Claim Based on Copilot**

17   OpenAI's second argument is that Plaintiffs fail to state a claim based on Copilot. In

18   OpenAI's view, "Plaintiffs' theory is based on actions by other Defendants and does not give rise

19   to a breach of contract claim against OpenAI." ECF No. 219 at 15. Plaintiffs respond that their

20   breach of contract claim must stand, as they have alleged that Copilot and Codex are "related,"

21   and that "Copilot is a joint venture relationship" between GitHub and OpenAI. ECF No. 235 at

22   21–22 (citing ECF No. 201 ¶ 59).

23   "A joint venture is 'an undertaking by two or more persons jointly to carry out a single

24   business enterprise for profit.'" *Forest v. Equitable Life Assurance Soc'y of U.S*., No. C99-5173

25   SI, 2001 WL 1338809, at *5 (N.D. Cal. June 12, 2001) (quoting *Nelson v. Abraham*, 29 Cal. 2d

26   745, 749 (1947)). "The elements necessary for a joint venture are: (1) an intent to become

27   partners; (2) a community of interest in the undertaking; (3) an understanding to share profits and

28   losses; and (4) equal authority and right to direct and control the conduct of all co-venturers with

1   respect to the joint venture." *Forest*, 2001 WL 1338809, at *5. Plaintiffs allege that "Copilot

2   requires Codex to function," and that "[e]ach [Defendant] acted as the principal, agent, or joint

3   venture of, or for other Defendants with respect to the acts, violations, and common course of

4   conduct alleged herein." ECF No. 201 ¶¶ 26, 45. Additionally, they aver that "Codex is a

5   standalone product released by OpenAI that also 'powers GitHub Copilot, which [OpenAI] built

6   and launched in partnership with GitHub'" and that "'GitHub Copilot uses the OpenAI Codex to

7   suggest code and entire functions in real-time, right from your editor.'" *Id.* ¶ 59. Finally, they

8   allege that "[t]he profits attributable to Defendants' violation . . . include the revenue from:

9   Copilot subscription fees, sales of or subscriptions to Defendants' Copilot-related products and/or

10   services that are used to run Copilot, hosting Copilot on Azure, and any other of Defendants'

11   products." *Id.* ¶ 225.[6]

12   OpenAI responds that "Copilot is not a joint venture because OpenAI lacks 'equal

13   authority and right to direct and control the conduct' of GitHub with respect to Copilot, as

14   demonstrated by the fact that GitHub alone has released features altering Copilot's outputs." ECF

15   No. 244 at 15–16 (citing ECF No. 201 ¶¶ 145–157). True, Plaintiffs' allegations concerning the

16   duplication-detection feature state that "GitHub Copilot now includes an option to either allow or

17   block code completion suggestions that match publicly available code." ECF No. 201 ¶ 145. But

18   this does not dispel Plaintiffs' other allegations that "Copilot requires Codex to function," and that

19   "Codex . . . powers GitHub Copilot." *Id.* ¶ 59. Accordingly, the Court declines to dismiss

20   Plaintiffs' breach of contract claim on this ground.

4.   Conditions Versus Covenants

22   Finally, Defendants contend that even if they violated the attribution and notice terms of

23   the Doe Licenses, those breaches were of conditions that sound in copyright law, and therefore

24   Plaintiffs' claims for breach of contract must be dismissed.[7]

25

---

26   [6] While this allegation is specific to Plaintiffs' DMCA claim, the Court finds that it sufficiently
     alleges that there is "an understanding to share profits and losses" amongst Defendants.

27   [7] Although it appears that Plaintiffs forgot to attach the full text of the Doe Licenses to the SAC,
     as they did for their first two complaints, the Court will nonetheless incorporate these licenses by

28   reference. Plaintiffs refer to the Doe Licenses throughout the SAC and assert a claim for breach of
     contract based on the content of those licenses. *United States v. Ritchie*, 342 F.3d 903, 908 (9th

9

1   Before delving into the specifics of this argument, a brief refresher on the intersection

2 between covenants, conditions, and copyright law is in order. "Generally, a copyright owner who

3 grants a nonexclusive license to use his copyrighted material waives his right to sue the licensee

4 for copyright infringement and can sue only for breach of contract." *Jacobsen v. Katzer*, 535 F.3d

5 1373, 1380 (Fed. Cir. 2008) (quoting *Sun Microsystems, Inc., v. Microsoft Corp*., 188 F.3d 1115,

6 1121 (9th Cir. 1999); *Graham v. James*, 144 F.3d 229, 236 (2d Cir. 1998)). "If, however, a

7 license is limited in scope and the licensee acts outside the scope, the licensor can bring an action

8 for copyright infringement." *Jacobsen*, 535 F.3d at 1380 (citing *S.O.S., Inc. v. Payday, Inc*., 886

9 F.2d 1081, 1087 (9th Cir. 1989); *Nimmer on Copyright* § 1015[A]). "[C]ontractual terms that

10 limit a license's scope [are] 'conditions,' the breach of which constitute copyright infringement."

11 *MDY Indus., LLC v. Blizzard Ent., Inc*., 629 F.3d 928, 939 (9th Cir. 2010), *as amended on denial*

12 *of reh'g* (Feb. 17, 2011), *opinion amended and superseded on denial of reh'g*, No. 09-15932, 2011

13 WL 538748 (9th Cir. Feb. 17, 2011). "[A]ll other license terms [are] 'covenants,' the breach of

14 which is actionable only under contract law." *Id*. "Conditions precedent are disfavored and will

15 not be read into a contract unless required by plain, unambiguous language." *Effects Associates,*

16 *Inc. v. Cohen*, 908 F.2d 555, 559 n.7 (9th Cir. 1990).

17   While OpenAI is likely correct that the attribution and notice terms in the Doe Licenses at

18 issue are conditions, this does not impede Plaintiffs' ability to bring a breach of contract claim.

19 *Patry on Copyright* is instructive: "[i]t is common for courts to say that if there is a material

20 breach of a condition of the license, the copyright owner has the *option* of suing for copyright

21 infringement *or* breach of contract, but if there is a violation of a covenant, only a breach-of-

22 contract claim will lie." 5 *Patry on Copyright* § 17:43 (emphasis added); *see also* 3 *Nimmer on*

23 *Copyright* § 10.15 (emphasis added) ("If the grantee's violation consists of a failure to satisfy a

24 condition to the grant (as distinguished from a breach of a covenant), it follows that . . . the

25

26 Cir. 2003) (holding that documents "may be incorporated by reference into a complaint if the
plaintiff refers extensively to the document or the document forms the basis of the plaintiff's
27 claim."). Relevant here, Plaintiffs argue that Defendants breached six open-source licenses: the
MIT License (ECF No. 98-1 at 58), GNU General Public License version 2.0 (*id*. at 28–34), GNU
28 General Public License version 3.0 (*id*. at 35–48), GNU Affero General Public License 3.0 (*id*. at
16–27), 3-Clause BSD License (*id*. at 9), and Apache License 2.0 (*id*. at 2–6).

1   grantee's conduct *may* constitute copyright infringement."); *Costello Pub. Co. v. Rotelle*, 670 F.2d

2   1035, 1041 (D.C. Cir. 1981) (citing 3 *Nimmer on Copyright* § 10.15) ("if Talbot Press failed to

3   satisfy a condition to the license, any use by the licensee or its assignee would constitute an

4   infringement of copyright and defendant-intervenors could elect to pursue a remedy for

5   infringement rather than breach of contract."); *Sohm v. Scholastic Inc.*, 959 F.3d 39, 46 (2d Cir.

6   2020) (quotation marks and alteration omitted) ("If a license is limited in scope and the licensee

7   acts outside the scope, the licensor can bring an action for copyright infringement."); *Sun*

8   *Microsystems*, 188 F.3d at 1121 (if the licensee fails to satisfy the condition and "acts outside the

9   scope [of the license], the licensor can bring an action for copyright infringement."). Accordingly,

10   the Court declines to read in a requirement that a plaintiff *must* bring suit for copyright

11   infringement in the event of a breach of condition.

12   Finally, OpenAI contends that *Jacobsen* "is relevant because it found that terms similar to

13   those in the Doe Licenses were conditions to the license grant (giving rise to a copyright claim),

14   and not covenants undertaken by the licensee (giving rise to a contract claim)." ECF No. 244 at

15   19. In OpenAI's view, "[t]he same analysis applies whether the plaintiff ultimately brings a

16   copyright or contract claim." *Id.* The Court disagrees. Although the language of the licenses in

17   *Jacobsen* is fairly analogous to the language of the Doe Licenses, this is immaterial to whether

18   Plaintiffs can bring a breach of contract claim. As evidenced by the treatises and caselaw above,

19   suing for copyright infringement is not the exclusive avenue a plaintiff must pursue in the event of

20   a breach of a condition of a license—it is simply one option a plaintiff may elect.

21   Therefore, the Court concludes that Plaintiffs have alleged a breach of contract claim for

22   violation of open-source licenses.

23   **C.**   **Unjust Enrichment and Punitive Damage Requests**

24   **1.**   **Unjust Enrichment**

25   Finally, the Court addresses Defendant GitHub's argument that Plaintiffs' request for

26   monetary relief in the form of unjust enrichment, as well as their request for punitive damages,

27   should be denied. ECF No. 214-2 at 24–25. Agreeing with GitHub on both fronts, the Court

28   dismisses Plaintiffs' requests for unjust enrichment and punitive damages.

1   Beginning with unjust enrichment, "[u]njust enrichment is not a cause of action [] or even

2 a remedy, but rather a general principle, underlying various legal doctrines and remedies. It is

3 synonymous with restitution." *McBride v. Boughton*, 123 Cal. App. 4th 379, 387 (2004) (internal

4 quotations and citations omitted). "There are several potential bases for a cause of action seeking

5 restitution. For example, restitution may be awarded in lieu of breach of contract damages when

6 the parties had an express contract, but it was procured by fraud or is unenforceable or ineffective

7 for some reason. Alternatively, restitution may be awarded where the defendant obtained a benefit

8 from the plaintiff by fraud, duress, conversion, or similar conduct. In such cases, the plaintiff may

9 choose not to sue in tort, but instead to seek restitution on a quasi-contract theory." *Durell v.*

10 *Sharp Healthcare*, 183 Cal. App. 4th 1350, 1370 (2010) (internal quotations and citations

11 omitted).

12   GitHub argues that Plaintiffs' "request for unjust enrichment is insupportable under

13 California law." ECF No. 242-1 at 18. GitHub avers that "although unjust enrichment is

14 sometimes a 'theory underlying a claim that a defendant has been unjustly conferred a benefit,' a

15 plaintiff must [also] plead 'mistake, fraud, coercion, or request.'" ECF No. 214-2 at 25 (quoting

16 *Astiana v. Hain Celestial Grp., Inc.*, 783 F.3d 753, 762 (9th Cir. 2015)). California law supports

17 this position: absent an exception, "a quasi-contract action for unjust enrichment does not lie

18 where, as here, express binding agreements exist and define the parties' rights." *California Med.*

19 *Ass'n, Inc. v. Aetna U.S. Healthcare of California, Inc.*, 94 Cal. App. 4th 151, 172 (2001). The

20 Court agrees with GitHub that Plaintiffs' breach of contract claims do not contain any allegations

21 of mistake, fraud, coercion, or request. Accordingly, unjust enrichment damages are not

22 available.[8]

23   In response, Plaintiffs point to the Restatement (Third) of Restitution and Unjust

24 Enrichment § 39, which provides that where "a deliberate breach of contract results in profit to the

25

26 _____

27 [8] To the extent Plaintiffs' request for unjust enrichment monetary relief stems from their
standalone claim of unjust enrichment that was previously dismissed with prejudice, the Court
agrees with GitHub that such a request is improper and must be stricken. *See Santa Clara Valley*

28 *Water Dist. v. Olin Corp.*, No. 07-cv-03756-RMW, 2007 WL 2890390, at *5 (N.D. Cal. Sept. 28,
2007) ("Improper prayers for relief are proper subjects for a motion to strike.").

1   defaulting promisor and the available damage remedy affords inadequate protection to the

2   romise's contractual entitlement, the   romise has a claim to restitution of the profit realized by

3   the promisor as a result of the breach." Restatement (Third) of Restitution and Unjust Enrichment,

4   § 39 (2011); *see* ECF No. 234 at 23. Plaintiffs, however, cite to no California authority (and the

5   Court is aware of none) that adopts Section 39. *See AcryliCon USA, LLC v. Silikal GmbH*, 985

6   F.3d 1350, 1372 (11th Cir. 2021) (rejecting plaintiff's reliance on Section 39 in asserting that, as a

7   remedy for its breach-of-contract claims, it was entitled to disgorgement of defendant's profits, as

8   Georgia law did not follow Section 39 in permitting disgorgement as a remedy for claims

9   sounding in contract.). Further, Plaintiffs contend that 55 Cal. Jur. 3d Restitution § 2 supports

10   their position, as it states that "[a] party to an express contract can assert a claim for restitution

11   based on unjust enrichment by alleging in that cause of action that the express contract is void or

12   was rescinded." ECF No. 234 at 24. But critically, Plaintiffs do not assert that the contracts at

13   issue are either void or rescinded.

14   The caselaw Plaintiffs provide is similarly unavailing. In *MSC.Software Corp. v. Heroux-*

15   *Devtek Inc.*, No. 8:19-cv-01987-SB-(DFMx), 2021 WL 9696752, at *2 (C.D. Cal. Sept. 16, 2021),

16   the court concluded that "contrary to Defendant's assertion, Plaintiff's request for restitution may

17   still be tried by a jury." However, in reaching that conclusion the *MSC* court relied on caselaw

18   concerning quasi-contract. *See Welborne v. Ryman-Carroll Found.*, 22 Cal. App. 5th 719, 725

19   (2018) ("A cause of action for quasi-contract invokes consideration of equitable principles, rather

20   than of contract . . . . In applying the principles of unjust enrichment, . . . a plaintiff is entitled to

21   restitution of the amount at issue."). The other cases Plaintiffs cite fall prey to the same issue. *See*

22   *Hernandez v. Lopez*, 180 Cal. App. 4th 932, 938–39 (2009) (allowing plaintiffs to recover for

23   unjust enrichment on a quasi-contract theory); *Alkayali v. Hoed*, No. 18- cv-777, 2018 WL

24   3425980, at *6 (S.D. Cal. July 16, 2018) (citing cases relying on quasi-contract).

25   One final point bears mention. The Court conducted additional research concerning

26   whether "a defendant's unjust enrichment can satisfy the 'damages' element of a breach of

27   contract claim, such that disgorgement is a proper remedy." *Foster Poultry Farms, Inc. v.*

28   *SunTrust Bank*, 377 Fed. Appx. 665, 669 (9th Cir. 2010). In *Foster Poultry Farms, Inc.*, the Ninth

1   held that "[u]nder California law, disgorgement of improperly obtained profits can be an

2   appropriate remedy for breach of a contract . . . ." *Id.* (citing *Ajaxo Inc. v. E*Trade Group, Inc.*,

3   135 Cal. App. 4th 21, 56 (2005)). The Court declines to follow the reasoning of this non-binding

4   case. In *Ajaxo Inc.*—the sole California case upon which *Foster Poultry Farms, Inc.*, relies—the

5   plaintiff established that the defendant provided protected information to a competitor, and in

6   doing so, violated the terms of their NDA. *Ajaxo Inc.*, 135 Cal. App. 4th at 55–56. As damages

7   for the breach of the NDA, the plaintiff was awarded unjust enrichment damages. *Id.* However,

8   that was not because the *Ajaxo Inc.* court determined that, as a matter of California law, unjust

9   enrichment damages were generally recoverable for a breach of contract. Rather, the NDA in

10   *Ajaxo Inc.* expressly "allow[ed] for an equitable remedy in addition to 'whatever remedies it might

11   have at law.'" *Id.* at 58. Put differently, the unjust enrichment measure of damages was explicitly

12   written into the parties' contract. *See Adcor Indus., Inc. v. Beretta U.S.A. Corp.*, 248 A.3d 1137,

13   1150 (2021) (concluding that the *Ajaxo Inc* court determined that unjust enrichment damages

14   applied because they were "grounded in the parties' contract," not because "unjust enrichment

15   damages, as a matter of California law, were generally recoverable for a breach of NDA."). To

16   this Court's knowledge, absent rare circumstances,[9] California courts do not hold that a party may

17   recover restitution of a defendant's unjust enrichment as a remedy where valid contractual

18   agreements exist and define the parties' rights.[10]

19   In sum, Plaintiffs' claims do not support they remedy they seek. Plaintiffs have failed to

20   establish, as a matter of law, that restitution for any unjust enrichment is available as a measure of

21   Plaintiffs' damages for their breach of contract claims.

22   **2.   Punitive Damages**

23   Turning to GitHub's arguments concerning punitive damages, the Court agrees that these

24

25   [9] *Dunkin v. Boskey*, 82 Cal. App. 4th 171 (2000), allowed recovery for breach of contract on an
     unjust enrichment theory. That case, however, involved a unique set of facts concerning an
26   artificial insemination contract involving an unmarried couple. After the child was born, the
     mother began to deny the male partner his rights under the contract. The court concluded that the
27   male partner could not recover under breach of contract for loss of the relationship with the child,
     but it held he could recover "special damages for readily ascertainable economic loss under an
28   unjust enrichment theory." *Id.* at 195.
     [10] *Foster Poultry Farms, Inc.* has also not been cited by any other Ninth Circuit decision.

claims must be dismissed. For one, Plaintiffs' opposition fails to address GitHub's motion to

dismiss Plaintiff's punitive damages claim. This alone warrants dismissal of Plaintiffs' claim for

punitive damages. *See Moore v. Apple, Inc.*, 73 F. Supp. 3d 1191, 1205 (N.D. Cal. 2014) (finding

that a failure in an opposition to address arguments raised in a motion to dismiss "constitutes

abandonment of the claim," which results in dismissal with prejudice). But regardless, the sole

remaining causes of action are for breach of contract, for which punitive damages are generally

not recoverable. *Harris v. Atl. Richfield Co.*, 14 Cal. App. 4th 70, 77 (1993) ("As a general rule,

California law does not authorize the award of general or punitive damages for breach of a

commercial contract."); *R Power Biofuels, LLC v. Chemex LLC*, No. 16-CV-00716-LHK, 2016

WL 6663002, at *21 (N.D. Cal. Nov. 11, 2016) ("the parties agree that punitive damages are only

available for contract claims if the breaches of contract are also tortious.").

CONCLUSION

In sum, the Court dismisses Plaintiffs' Section 1202(b) claim, this time with prejudice.

The Court declines to dismiss Plaintiffs' claim for breach of contract of open-source license

violations against all Defendants. Finally, the Court dismisses Plaintiffs' request for monetary

relief in the form of unjust enrichment, as well as Plaintiffs' request for punitive damages.
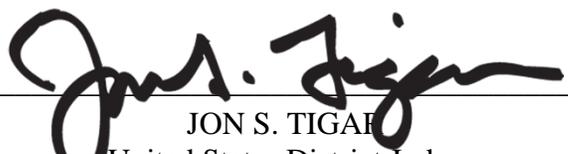
IT IS SO ORDERED.

Dated: June 24, 2024

_____
JON S. TIGAR
United States District Judge

EXHIBIT B

1

2

3

4   UNITED STATES DISTRICT COURT

5   NORTHERN DISTRICT OF CALIFORNIA

6

7   J. DOE 1, et al.,   Case No. 22-cv-06823-JST

8   Plaintiffs,

9   v.   **ORDER GRANTING MOTION TO
                                                 CERTIFY ORDER FOR
10   GITHUB, INC., et al.,   INTERLOCUTORY APPEAL AND
                                                 MOTION TO STAY PENDING
                    Defendants.   APPEAL**
11
                                                 Re: ECF No. 268
12

13   Before the Court is Plaintiffs' motion to certify the Court's June 24, 2024 order dismissing

14   their Section 1202(b) claims. ECF No. 253. The Court will grant the motion.

15   The final judgement rule ordinarily provides that courts of appeal shall have jurisdiction

16   only over "final decisions of the district courts of the United States." 28 U.S.C. § 1291. However,

17   "[w]hen a district judge, in making in a civil action an order not otherwise appealable under this

18   section, shall be of the opinion that such order involves a controlling question of law as to which

19   there is substantial ground for difference of opinion and that an immediate appeal from the order

20   may materially advance the ultimate termination of the litigation, he shall so state in writing in

21   such order." 28 U.S.C. § 1292(b). "The Court of Appeals which would have jurisdiction of an

22   appeal of such action may thereupon, in its discretion, permit an appeal to be taken from such

23   order." *Id.*

24   "Certification under § 1292(b) requires the district court to expressly find in writing that all

25   three § 1292(b) requirements are met." *Couch v. Telescope Inc.*, 611 F.3d 629, 633 (9th Cir.

26   2010). "Section 1292(b) is a departure from the normal rule that only final judgements are

27   appealable, and therefore must be construed narrowly." *James v. Price Stern Sloan, Inc.*, 283 F.3d

28   1064, 1067 n.6 (9th Cir. 2002). To that end, "section 1292(b) is to be applied sparingly and only

*United States District Court
Northern District of California*

1   in exceptional cases." *In re Cement Antitrust Litig. (MDL No. 296)*, 673 F.2d 1020, 1027 (9th Cir.

2   1981), *aff'd sub nom. Arizona v. Ash Grove Cement Co.*, 459 U.S. 1190 (1983).

3   Plaintiffs have satisfied the requirements of 28 U.S.C. § 1292(b). Accordingly, Plaintiffs'

4   motion will be granted.

5   First, the Court's dismissal of Plaintiffs' Section 1202(b) claims involves a "controlling

6   question of law". 28 U.S.C. § 1292(b). A "controlling" question of law may only be found in

7   "exceptional situations in which allowing an interlocutory appeal would avoid protracted and

8   expensive litigation." *In re Cement Antitrust Litig.*, 673 F.2d at 1026. A question of law is

9   controlling if the "resolution of the issue on appeal could materially affect the outcome of

10   litigation in the district court." *Id.* at 1026. A controlling question of law "generally is a purely

11   legal one that can be resolved quickly without delving into a particular case's facts." *Henley v.*

12   *Jacobs*, No. C 18-2244 SBA, 2019 WL 8333448, at *2 (N.D. Cal. Oct. 25, 2019). The question of

13   law raised by Plaintiffs is whether Sections 1202(b)(1) and (b)(3) of the DMCA impose an

14   identicality requirement. This is a purely legal question of statutory interpretation. *See In re*

15   *Google Inc. St. View Elec. Commc'ns Litig.*, No. C 10-MD-02184 JW, 2011 WL 13257346, at *1

16   (N.D. Cal. July 18, 2011) (finding a "controlling question of law" suitable for interlocutory appeal

17   regarding a "novel question of statutory interpretation'). Moreover, if this issue were to be

18   decided in Plaintiffs' favor they would be able to proceed with their DMCA claims, which counsel

19   claims are the "heart of their class case." ECF No. 268 at 8.

20   Second, the Court finds there is substantial ground for difference of opinion on the

21   question at issue. 28 U.S.C. § 1292(b). To determine whether there is a "substantial ground for

22   difference of opinion," courts examine "to what extent the controlling law is unclear." *Couch*, 611

23   F.3d at 633. "[A] party's strong disagreement with the Court's ruling is not sufficient for there to

24   be a 'substantial ground for difference.'" *Id.* (quotation marks omitted). A substantial ground for

25   difference of opinion may exist where "the circuits are in dispute on the question and the court of

26   appeals of the circuit has not spoken on the point, if complicated questions arise under foreign

27   law, or if novel and difficult questions of first impression are presented." *Id.* (quoting 3 Federal

28   Procedure, Lawyers Edition § 3:212 (2010) (footnotes omitted)). To this Court's knowledge no

United States District Court
Northern District of California

1   court of appeal has ruled on this issue, and district courts have reached differing conclusions. *See,*

2   *e.g.*, *Anderson v. Stability AI Ltd.*, No. 23-cv-00201-WHO, 2024 WL 3823234, at *8 (N.D. Cal.

3   Aug. 12, 2024) (finding DMCA Section 1202(b) claims have an identicality requirement); *Beijing*

4   *Meishe Network Tech. Co., Ltd., v. TikTok Inc.*, No. 23-cv-06012-SI, 2024 WL 3522196, at *9

5   (N.D. Cal. July 23, 2024) (declining to dismiss on identicality grounds); *ADR Int'l Ltd. v. Inst. for*

6   *Supply Mgmt. Inc.*, 667 F. Supp. 3d 411, 427 (S.D. Tex. 2023) ("Based on the plain wording of the

7   statute, the Court is not persuaded that the DMCA includes an 'identical copy' requirement.");

8   *Kirk Kara Corp v. W. Stone & Metal Corp.*, No. CV 20-1931-DMG (Ex), 2020 WL 5991503

9   (C.D. Cal. Aug. 14, 2020) (adopting an "identicality" standard); *Oracle Int'l Corp. v. Rimini St.,*

10   *Inc.*, No. 2:14-cv-01699-MMD-DJA, 2023 WL 4706127, at *82 (D. Nev. July 24, 2023) ("The

11   Court also rejects Rimini's argument . . . that a work that removes copyright management

12   information must be an exact copy of the original work. This construction of the DMCA would

13   weaken the statute's intended protections for copy right holders."). "One of the best indications

14   that there are substantial grounds for disagreement on a question of law is that other courts have,

15   in fact, disagreed." *Rollins v. Dignity Health*, No. 13-cv-01450-TEH, 2014 WL 6693891, at *3

16   (N.D. Cal. Nov. 26, 2014). Given this split in authority the Court finds there are substantial

17   grounds for disagreement on this issue.

18   Finally, the immediate appeal from the order is likely to materially advance the ultimate

19   outcome of the litigation. 28 U.S.C. § 1292(b). Courts within this Circuit have held "that

20   resolution of a question materially advances the termination of litigation if it 'facilitate[s]

21   disposition of the action by getting a final decision on a controlling legal issue sooner, rather than

22   later [in order to] save the courts and the litigants unnecessary trouble and expense.'" *Finder v.*

23   *Leprino Foods Co.*, No. 13-cv-02059-AWI-BAM, 2016 WL 4095833, at *4 (E.D. Cal. Aug. 1,

24   2016) (citations omitted). The Ninth Circuit resolving this issue would materially advance not

25   only this particular case, but others in the Circuit raising the same issue. "Rather than litigating

26   the case to the finish [without the DMCA claims which] will be challenged on appeal, the Court

27   and the parties will benefit from definitive guidance from the Ninth Circuit at the outset, before

28   time and resources are invested." *J.B. v. G6 Hosp., LLC*, No. 19-cv-07848-HSG, 2021 WL

1 6621068, at *4 (N.D. Cal. Dec. 16, 2021).

2   Having concluded that it is appropriate to certify the order for interlocutory appeal, the

3 Court now addresses Plaintiffs' request for a stay pending appeal. When determining if a stay is

4 appropriate, courts apply a three-factor test derived from *Landis v. North American Co.*, 299 U.S.

5 248 (1936) and consider (1) "the possible damage which may result from the granting of a stay;

6 (2) the hardship or inequity which a party may suffer [if the case is allowed] to go forward; and (3)

7 the orderly course of justice measured in terms of the simplifying or complicating of issues, proof,

8 and questions of law which could be expected to result from a stay." *Kuang v. U.S. Dep't of Def.*,

9 No. 18-CV-03698-JST, 2019 WL 1597495, a *2 (N.D. Cal. Apr. 15, 2019) (quoting *Lockyer v.*

10 *Mirant Corp.*, 398 F.3d 1098, 1110 (9th Cir. 2005)).

11   With respect to the first *Landis* factor, Defendants do not contend any damage will result

12 from granting a stay or that they will face any hardship. The parties are still in the early stages of

13 discovery. Turning to the second factor, if the Court declines to issue a stay and the Ninth Circuit

14 revives Plaintiffs' DCMA claims, Plaintiffs may have to redo fact discovery, expert testimony,

15 and/or class certification. Accordingly, if the case were to proceed pending Ninth Circuit review,

16 "significant and potentially unnecessary" resources might be invested by both parties. *Gustavson*

17 *v. Mars, Inc.*, No. 13-cv-04537-LHK, 2014 WL 6986421, at *3 (N.D. Cal. Dec. 10, 2014).

18 Finally, the Court believes that "judicial economy will be best served" by staying this case as "the

19 Ninth Circuit's decision is likely to provide substantial guidance" that may "materially alter the

20 Court's decisions in the instant case." *Id.*

21   **CONCLUSION**

22   For the foregoing reasons the Court grants Plaintiffs' motion to seek interlocutory appeal.

23 The Court also grants Plaintiffs' request for a stay of trial court proceedings while the order

24 / / /

25 / / /

26 / / /

27 / / /

28 / / /

United States District Court
Northern District of California

1   certifying interlocutory appeal is reviewed by the Ninth Circuit. The parties are ordered to notify

2   the Court within 10 days of the receipt of a decision from the Ninth Circuit Court of Appeals.

3   **IT IS SO ORDERED.**

4   Dated: September 27, 2024

5   _____
      JON S. TIGAR
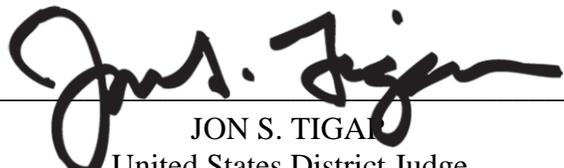
6   United States District Judge

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

United States District Court
Northern District of California

5

EXHIBIT C

Joseph R. Saveri (State Bar No. 130064)
Cadio Zirpoli (State Bar No. 179108)
Travis Manfredi (State Bar No. 281779)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1000
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
                czirpoli@saverilawfirm.com
                tmanfredi@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
Telephone: (323) 968-2632
Facsimile: (415) 395-9940
Email: mb@buttericklaw.com

*Counsel for Individual and Representative
Plaintiffs and the Proposed Class*

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN FRANCISCO DIVISION

J. DOE 1 and J. DOE 2, individually and on behalf of all others similarly situated,
 	
    Individual and Representative Plaintiffs,
 	
    v.
 	
GITHUB, INC., a Delaware corporation; MICROSOFT CORPORATION, a Washington corporation; OPENAI, INC., a Delaware nonprofit corporation; OPENAI, L.P., a Delaware limited partnership; OPENAI GP, L.L.C., a Delaware limited liability company; OPENAI STARTUP FUND GP I, L.L.C., a Delaware limited liability company; OPENAI STARTUP FUND I, L.P., a Delaware limited partnership; OPENAI STARTUP FUND MANAGEMENT, LLC, a Delaware limited liability company,
 	
    Defendants.
	Case No.
	
	COMPLAINT
	
	CLASS ACTION
	
	DEMAND FOR JURY TRIAL

TABLE OF CONTENTS

           CLASS ACTION COMPLAINT

Plaintiffs J. Doe 1 and J. Doe 2 ("Plaintiffs"), on behalf of themselves and all others similarly situated, bring this Class Action Complaint (the "Complaint") against Defendants GitHub, Inc.; Microsoft Corporation; OpenAI, Inc.; OpenAI, L.P.; OpenAI GP, L.L.C.; OpenAI Startup Fund GP I, L.L.C.; OpenAI Startup Fund I, L.P.; and OpenAI Startup Fund Management, LLC[1] for violation of the Digital Millennium Copyright Act, 17 U.S.C. §§ 1201–1205 (the "DMCA"); violation of the Lanham Act, 15 U.S.C. § 1125; violation of Unfair Competition law, *Cal. Bus. & Prof. Code* §§ 17200, *et seq.*; violation of the California Consumer Privacy Act, *Cal. Civ. Code* § 1798.150 (the "CCPA"); and Breach of Contract regarding the Suggested Licenses, GitHub's Privacy Statement, and GitHub's Terms of Service, *Cal. Bus. & Prof. Code* §§ 22575–22579, *Cal. Civ. Code* § 1798.150. Plaintiffs and the Class also bring this Complaint against Defendants for their Tortious Interference in Plaintiffs' Contractual Relationships; Fraud, and Negligence regarding handling of sensitive data.

I.   OVERVIEW: A BRAVE NEW WORLD OF SOFTWARE PIRACY

1.   Plaintiffs and the Class are owners of copyright interests in materials made available publicly on GitHub that are subject to various licenses containing conditions for use of those works (the "Licensed Materials."). All the licenses at issue here (the "Licenses") contain certain common terms (the "License Terms").

2.   "Artificial Intelligence" is referred to herein as "AI." AI is defined for the purposes of this Complaint as a computer program that algorithmically simulates human reasoning or inference, often using statistical methods. Machine Learning ("ML") is a subset of AI in which the behavior of the program is derived from studying a corpus of material called training data.

---

[1] GitHub, Inc. is referred to as "GitHub." Microsoft Corporation is referred to as "Microsoft." OpenAI, Inc.; OpenAI, L.P.; OpenAI GP, L.L.C.; OpenAI Startup Fund GP I, L.L.C.; OpenAI Startup Fund I, L.P.; and OpenAI Startup Fund Management, LLC are referred to collectively herein as "OpenAI." Collectively, GitHub, Inc., Microsoft Corporation, OpenAI, Inc.; OpenAI, L.P.; OpenAI GP, L.L.C.; OpenAI Startup Fund GP I, L.L.C.; OpenAI Startup Fund I, L.P.; and OpenAI Startup Fund Management, LLC are referred to herein as "Defendants."

3.   GitHub is a company founded in 2008 by a team of open-source enthusiasts. At the time, GitHub's stated goal was to support open-source development, especially by hosting open-source source code on the website github.com. Over the next 10 years, GitHub, based on these representations succeeded wildly, attracting nearly 25 million developers.

4.   Developers published Licensed Materials on GitHub pursuant to written Licenses. In particular, the most popular ones share a common term: use of the Licensed Materials requires some form of *attribution*, usually by, among other things, including a copy of the license along with the name and copyright notice of the original author.

5.   On October 26, 2018, Microsoft acquired GitHub for $7.5 billion. Though some members of the open-source community were skeptical of this union, Microsoft repeated one mantra throughout: "Microsoft Loves Open Source". For the first few years, Microsoft's representations seemed credible.

6.   Microsoft invested $1 billion in OpenAI LP in July 2019 at a $20 billion valuation. In 2020, Microsoft became exclusive licensee of OpenAI's GPT-3 language model—despite OpenAI's continued claims its products are meant to benefit "humanity" at large. In 2021, Microsoft began offering GPT-3 through its Azure cloud-computing platform. On October 20, 2022, it was reported that OpenAI "is in advanced talks to raise more funding from Microsoft" at that same $20 billion valuation. Copilot runs on Microsoft's Azure platform. Microsoft has used Copilot to promote Azure's processing power, particularly regarding AI.

7.   On information and belief, Microsoft obtained a partial ownership interest in OpenAI in exchange for its $1 billion investment. As OpenAI's largest investor and largest service provider—specifically in connection with Microsoft's Azure product—Microsoft exerts considerable control over OpenAI.

8.   In June 2021, GitHub and OpenAI launched Copilot, an AI-based product that promises to assist software coders by providing or filling in blocks of code using AI. GitHub charges Copilot users $10 per month or $100 per year for this service. Copilot ignores, violates, and removes the Licenses offered by thousands—possibly millions—of software developers, thereby accomplishing software piracy on an unprecedented scale. Copilot outputs text derived

1 from Plaintiffs' and the Class's Licensed Materials without adhering to the applicable License

2 Terms and applicable laws. Copilot's output is referred herein as "Output."

3   9.   On August 10, 2021, OpenAI debuted its Codex product, which converts natural

4 language into code and is integrated into Copilot. (Copilot and Codex can be called either AIs or

5 MLs. Herein they will be referred to as AIs unless a distinction is required.)

6   10.   Though Defendants have been cagey about what data was used to train the AI,[2]

7 they have conceded that the training data includes data in vast numbers of publicly accessible

8 repositories on GitHub,[3] which include and are limited by Licenses.

9   11.   Among other things, Defendants stripped Plaintiffs' and the Class's attribution,

10 copyright notice, and license terms from their code in violation of the Licenses and Plaintiffs' and

11 the Class's rights. Defendants used Copilot to distribute the now-anonymized code to Copilot

12 users as if it were created by Copilot.

13   12.   Copilot is run entirely on Microsoft's Azure cloud-computing platform.

14   13.   Copilot often simply reproduces code that can be traced back to open-source

15 repositories or open-source licensees. Contrary to and in violation of the Licenses, code

16 reproduced by Copilot *never* includes attributions to the underlying authors.

17   14.   GitHub and OpenAI have offered shifting accounts of the source and amount of

18 the code or other data used to train and operate Copilot. They have also offered shifting

19 justifications for why a commercial AI product like Copilot should be exempt from these license

20 requirements, often citing "fair use."

21   15.   It is not fair, permitted, or justified. On the contrary, Copilot's goal is to replace a

22 huge swath of open source by taking it and keeping it inside a GitHub-controlled paywall. It

23 violates the licenses that open-source programmers chose and monetizes their code despite

24 GitHub's pledge never to do so.

---

25 [2] "Training" an AI, as described in greater detail below, means feeding it large amounts of data

26 that it interprets using given criteria. Feedback is then given to it to fine-tune its Output until it
can provide Output with minimal errors.

27 [3] Repositories are containers for individual coding projects. They are where GitHub users upload

28 their code and where other users can find it. Most GitHub users have multiple repositories.

II.   JURISDICTION AND VENUE

16.   Plaintiffs bring this action on their own behalf as well as representatives of a Class of similarly situated individuals and entities. They seek to recover injunctive relief and damages as a result and consequence of Defendants' unlawful conduct.

17.   Jurisdiction and venue are proper in this judicial district pursuant to Defendants' violation of the Digital Millennium Copyright Act, 17 U.S.C. §§ 1201–1205 (the "DMCA"); Reverse Passing Off, Unjust Enrichment, and Unfair Competition under the Lanham Act, 15 U.S.C. § 1125; and because a substantial part of the events giving rise to Plaintiff's claims occurred in this District, Plaintiff J. Doe 1 resides in California, a substantial portion of the affected interstate trade and commerce was carried out in this District, and three or more of the Defendants reside in this District and/or are licensed to do business in this District. Each Defendant has transacted business, maintained substantial contacts, and/or committed overt acts in furtherance of the illegal scheme and conspiracy throughout the United States, including in this District. Defendants' conduct has had the intended and foreseeable effect of causing injury to persons residing in, located in, or doing business throughout the United States, including in this District.

III.   INTRADISTRICT ASSIGNMENT

18.   Pursuant to Civil Local Rule 3.2 (c) and (e), assignment of this case to the San Francisco Division of the United States District Court for the Northern District of California is proper because a substantial amount of the development of the Copilot product as well as of the interstate trade and commerce involved and affected by Defendants' conduct giving rise to the claims herein occurred in this Division. Furthermore, Defendants GitHub and all the OpenAI entities are headquartered within this Division.

IV.   PARTIES

PLAINTIFFS

19.   Plaintiff J. Doe 1 is a resident of the State of California. Plaintiff Doe 1 published Licensed Materials they owned a copyright interest in to at least one GitHub repository under one of the Suggested Licenses. Specifically, Doe 1 has published Licensed Materials they claim a

copyright interest in under the following Suggested Licenses: MIT License and GNU General

Public License version 3.0. Plaintiff was, and continues to be, injured during the Class Period as a

result of Defendants' unlawful conduct alleged herein.

20.   Plaintiff J. Doe 2 is a resident of the State of Illinois. Plaintiff Doe 2 published

Licensed Materials they owned a copyright interest in to at least one GitHub repository under

one of the Suggested Licenses. Specifically, Doe 2 has published Licensed Materials they claim a

copyright interest in under the following Suggested Licenses: MIT License; GNU General Public

License version 3.0; GNU Affero General Public License version 3.0; The 3-Clause BSD

License; and Apache License 2.0. Plaintiff was, and continues to be, injured during the Class

Period as a result of Defendants' unlawful conduct alleged herein.

DEFENDANTS

21.   Defendant GitHub, Inc. is a Delaware corporation with its principal place of

business located at 88 Colin P Kelly Jr Street, San Francisco, CA 94107. GitHub sells, markets,

and distributes Copilot throughout the internet and other sales channels throughout the United

States, including in this District. GitHub released Copilot on a limited "technical preview" basis

on June 29, 2021. On June 21, 2022, Copilot was released to the public as a subscription-based

service for individual developers. GitHub is a party to the unlawful conduct alleged herein.

22.   Defendant Microsoft Corporation is a Washington corporation with its principal

place of business located at One Microsoft Way, Redmond, Washington 98052. Microsoft

announced its acquisition of Defendant GitHub, Inc. on June 4, 2018. On October 26, 2018,

Microsoft finalized its acquisition of GitHub. Microsoft owns and operates GitHub. Through its

corporate ownership, control of the GitHub Board of Directors, active management, and other

means, Microsoft sells, markets, and distributes Copilot. Microsoft is a party to the unlawful

conduct alleged herein.

23.   Defendant OpenAI, Inc. is a Delaware nonprofit corporation with its principal

place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI, Inc. is a party to

the unlawful conduct alleged herein. It—along with OpenAI, L.P.—programed, trained, and

maintains Codex, which infringes all the same rights at Copilot and is also an integral piece of

1   Copilot. Copilot requires Codex to function. OpenAI, Inc. is a party to the unlawful conduct

2   alleged herein. OpenAI, Inc. founded, owns, and exercises control over all the other OpenAI

3   entities, including those set forth in Paragraphs 24–28.

4   24.   Defendant OpenAI, L.P. is a Delaware limited partnership with its principal place

5   of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI, L.P. is a party to the

6   unlawful conduct alleged herein. Its primary activity is research and technology. OpenAI, L.P. is a

7   wholly owned subsidiary of OpenAI, Inc. that is operated for profit. OpenAI, L.P. is the OpenAI

8   entity that co-created Copilot and offers it jointly with GitHub. OpenAI's revenue, including

9   revenue from Copilot, is received by OpenAI, L.P. OpenAI, Inc. controls OpenAI, L.P. directly

10   and through the other OpenAI entities.

11   25.   Defendant OpenAI GP, L.L.C. ("OpenAI GP") is a Delaware limited liability

12   company with its principal place of business located at 3180 18th Street, San Francisco, CA

13   94110. OpenAI GP is the general partner of OpenAI, L.P. OpenAI GP manages and operates the

14   day-to-day business and affairs of OpenAI, L.P. OpenAI GP is liable for the debts, liabilities and

15   obligations of OpenAI, L.P., including litigation and judgments. OpenAI GP is a party to the

16   unlawful conduct alleged herein. Its primary activity is research and technology. OpenAI GP is

17   the general partner of OpenAI, L.P. OpenAI GP was aware of the unlawful conduct alleged herein

18   and exercised control over OpenAI, L.P. throughout the Class Period. OpenAI, Inc. directly

19   controls OpenAI GP.

20   26.   Defendant OpenAI Startup Fund I, L.P. ("OpenAI Startup Fund I") is a Delaware

21   limited partnership with its principal place of business located at 3180 18th Street, San Francisco,

22   CA 94110. OpenAI Startup Fund I was instrumental in the foundation of OpenAI, L.P., including

23   the creation of its business strategy and providing initial funding. Through participation in

24   OpenAI Startup Fund I, certain entities and individuals obtained an ownership interest in

25   OpenAI, L.P. Plaintiffs are informed and believed, and on that basis allege that OpenAI Startup

26   Fund I participated in the organization and operation of OpenAI, L.P. OpenAI Startup Fund I is a

27   party to the unlawful conduct alleged herein. OpenAI Startup Fund I was aware of the unlawful

28   conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period.

27.   Defendant OpenAI Startup Fund GP I, L.L.C. ("OpenAI Startup Fund GP I") is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI Startup Fund GP I is the general partner of OpenAI Startup Fund I. OpenAI Startup Fund GP I manages and operates the day-to-day business and affairs of OpenAI Startup Fund I. OpenAI Startup Fund GP I is liable for the debts, liabilities and obligations of OpenAI Startup Fund I, including litigation and judgments. OpenAI Startup Fund GP I was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period. OpenAI Startup Fund GP I is a party to the unlawful conduct alleged herein. Sam Altman, co-founder, CEO, and Board member of OpenAI, Inc. is the Manager of OpenAI Startup Fund GP I. OpenAI Startup Fund GP I is the General Partner of OpenAI Startup Fund I, L.P.

28.   Defendant OpenAI Startup Fund Management, LLC ("OpenAI Startup Fund Management") is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI Startup Fund Management is a party to the unlawful conduct alleged herein. OpenAI Startup Fund Management was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period.

V.   AGENTS AND CO-CONSPIRATORS

29.   The unlawful acts alleged against the Defendants in this class action complaint were authorized, ordered, or performed by the Defendants' respective officers, agents, employees, representatives, or shareholders while actively engaged in the management, direction, or control of the Defendants' businesses or affairs.

30.   The Defendants' agents operated under the explicit and apparent authority of their principals.

31.   Each Defendant, and its subsidiaries, affiliates and agents operated as a single unified entity.

1   32.   Various persons and/or firms not named as Defendants herein may have

2   participated as coconspirators in the violations alleged herein and may have performed acts and

3   made statements in furtherance thereof.

4   33.   Each acted as the principal, agent, or joint venture of, or for other Defendants with

5   respect to the acts, violations, and common course of conduct alleged herein.

6   ## VI.   CLASS ALLEGATIONS

7   **A.   Class Definitions**

8   34.   Plaintiffs bring this action for damages and injunctive relief on behalf of

9   themselves and all others similarly situated as a class action pursuant to Rules 23(a), 23(b)(2), and

10   23(b)(3) of the Federal Rules of Civil Procedure, on behalf of the following Classes:

11   **"Injunctive Relief Class" under Rule 23(b)(2):**

12   All persons or entities domiciled in the United States that, (1)
     owned an interest in at least one US copyright in any work; (2)
13   offered that work under one of GitHub's Suggested Licenses[4]; and
     (3) stored Licensed Materials in any public GitHub repositories at
14   any time between January 1, 2015 and the present (the "Class
     Period").

15

16

17

18

19   ───────────────────────

[4] When a GitHub user creates a new repository, they have the option of selecting one of thirteen
20   licenses from a dropdown menu to apply to the contents of that repository. (They can also apply a
     different license later, or no license.) The Creative Commons Zero v1.0 Universal and the
21   Unlicense donate the covered work to the public domain and/or otherwise waive all copyrights
     and related rights. Because they do not contain the necessary provisions nor do they even allow
22   the owner to make copyright claims in most circumstances, they are not included in the Class
     Definition. We refer to the remaining eleven options as the "Suggested Licenses," which are: (1)
23   Apache License 2.0 ("Apache 2.0"); (2) GNU General Public License version 3 ("GPL-3.0");
     (3) MIT License ("MIT"); (4) The 2-Clause BSD License ("BSD 2"); (5) The 3-Clause BSD
24   License ("BSD 3"); (6) Boost Software License ("BSL-1.0"); (7) Eclipse Public License 2.0
     ("EPL-2.0"); (8) GNU Affero General Public License version 3 ("AGPL-3.0"); (9) GNU
25   General Public License version 2 ("GPL-2.0"); (10) GNU Lesser General Public License version
     2.1 ("LGPL-2.1"); and (11) Mozilla Public License 2.0 ("MPL-2.0"). These Suggested Licenses
26   each contain at least three common requirements for use of the Licensed Materials in a derivative
     work or copy: attribution to the owner of the Licensed Materials ("Attribution"), inclusion of a
27   copyright notice ("Copyright Notice"), and inclusion of the applicable Suggested License's text
     ("License Terms").
28

───────────────────────

8   CLASS ACTION COMPLAINT

"Damages Class" under Rule 23(b)(3):

> All persons or entities domiciled in the United States that, (1) owned an interest in at least one US copyright in any work; (2) offered that work under one of GitHub's Suggested Licenses; and (3) stored Licensed Materials in any public GitHub repositories at any time during the Class Period.

These "Class Definitions" specifically exclude the following person or entities:

a.   Any of the Defendants named herein;

b.   Any of the Defendants' co-conspirators;

c.   Any of Defendants' parent companies, subsidiaries, and affiliates;

d.   Any of Defendants' officers, directors, management, employees, subsidiaries, affiliates, or agents;

e.   All governmental entities; and

f.   The judges and chambers staff in this case, as well as any members of their immediate families.

B.   Numerosity

35.   Plaintiffs do not know the exact number of Class members, because such information is in the exclusive control of Defendants. Plaintiffs are informed and believe that there are at least thousands of Class members geographically dispersed throughout the United States such that joinder of all Class members in the prosecution of this action is impracticable.

C.   Typicality

36.   Plaintiffs' claims are typical of the claims of their fellow Class members because Plaintiffs and Class members all own code published under a License. Plaintiffs and the Class published work subject to a License to GitHub later used by Copilot. Plaintiffs and absent Class members were damaged by this and other wrongful conduct of Defendants as alleged herein. Damages and the other relief sought herein is common to all members of the Class.

D.   Commonality & Predominance

37.   Numerous questions of law or fact common to the entire Class arise from Defendants' conduct—including, but not limited to those identified below:

1.   **DMCA Violations**

- Whether Defendants' conduct violated the Class's rights under the DMCA when GitHub and OpenAI caused Codex and Copilot to ingest and distribute Licensed Materials without including any associated Attribution, Copyright Notice, or License Terms.

2.   **Contract-Related Conduct**

- Whether Defendants violated the Licenses governing use of the Licensed Materials by using them to train Copilot and for republishing those materials without appending the required Attribution, Copyright Notice, or License Terms.

- Whether Defendants interfered in contractual relations between the Class and the public regarding the Licensed Materials by concealing the License Terms.

- Whether GitHub committed Fraud when it promised not to sell or distribute Licensed Materials outside GitHub in the GitHub Terms of Service and Privacy Statement.

3.   **Unlawful-Competition Conduct**

- Whether Defendants passed-off the Licensed Materials as its own creation and/or Copilot's creation.

- Whether Defendants were unjustly enriched by the unlawful conduct alleged herein.

- Whether Defendants Copilot-related conduct constitutes Unfair Competition under California law.

4.   **Privacy Violations**

- Whether GitHub violated the Class's rights under the California Consumer Privacy Act ("CCPA"), the GitHub Privacy Statement, and/or the California Constitution by, *inter alia*, sharing the Class's sensitive personal information (or, in the alternative, by not addressing an ongoing data breach of which it is aware); creating a product that contains personal data GitHub cannot delete,

alter, nor share with the applicable Class member; and selling the Class's

personal data.

- Whether GitHub committed Negligence when it failed to stop a still-ongoing

data breach it was and continues to be aware of.

5.   Injunctive Relief

- Whether this Court should enjoin Defendants from engaging in the unlawful

conduct alleged herein. And what the scope of that injunction would be.

6.   Defenses

- Whether any affirmative defense excuses Defendants' conduct.

- Whether any statutes of limitation limit Plaintiffs' and the Class's potential for

recovery.

- Whether any applicable statutes of limitation should be tolled as a result of

Defendants' fraudulent concealment of their unlawful conduct.

38.   These and other questions of law and fact are common to the Class and

predominate over any questions affecting the Class members individually.

E.   Adequacy

39.   Plaintiffs will fairly and adequately represent the interests of the Class because

they have experienced the same harms as the Class and have no conflicts with any other members

of the Class. Furthermore, Plaintiffs have retained sophisticated and competent counsel ("Class

Counsel") who are experienced in prosecuting Federal and state class actions throughout the

United States and other complex litigation and have extensive experience advising clients and

litigating intellectual property, competition, contract, and privacy matters.

F.   Other Class Considerations

40.   Defendants have acted on grounds generally applicable to the Class, thereby

making final injunctive relief appropriate with respect to the Class as a whole.

41.   This class action is superior to alternatives, if any, for the fair and efficient

adjudication of this controversy. Prosecuting the claims pleaded herein as a class action will

1   eliminate the possibility of repetitive litigation. There will be no material difficulty in the

2   management of this action as a class action.

3   42.   The prosecution of separate actions by individual Class members would create the

4   risk of inconsistent or varying adjudications, establishing incompatible standards of conduct for

5   Defendants.

6   <div align="center">**VII.   FACTUAL ALLEGATIONS**</div>

7   **A.   Introduction**

8   43.   This class action against Defendants concerns an OpenAI product called Codex

9   and a GitHub product called Copilot.

10   44.   OpenAI began development of Codex sometime after OpenAI was founded in

11   December 2015 and released Codex on a limited basis in August 2021.

12   45.   GitHub began development of Copilot sometime in 2019, released it on a limited

13   basis in June 2021, and released it as a paid subscription service in June 2022.

14   46.   Codex and Copilot are assistive AI-based systems offered to software

15   programmers. These AI systems are each trained on a large corpus of publicly accessible software

16   code and other materials, including all the Licensed Materials. Defendants describe these

17   products as making coding easier by accepting a code "prompt" from a programmer and emitting

18   a possible completion of that code, which is referred to herein as Output, as set forth above.

19   Defendants claim Codex and Copilot do not retain copies of the materials they are trained on. In

20   practice, however, the Output is often a near-identical reproduction of code from the training

21   data.

22   47.   Codex and Copilot are related. Copilot is a joint venture between GitHub and

23   OpenAI. Codex is a standalone product released by OpenAI that also "powers GitHub Copilot,

24   which [OpenAI] built and launched in partnership with GitHub."[5] "GitHub Copilot uses the

25   OpenAI Codex to suggest code and entire functions in real-time, right from your editor."[6] The

26

27   [5] *See* https://openai.com/blog/openai-codex/.

28   [6] *See* https://github.com/features/copilot/.

version of Codex used to power Copilot is modified from OpenAI's standalone version.
Defendants have kept secret the details of Codex's modifications and its integration into or
interaction with Copilot.

B.   Codex Outputs Copyrighted Materials Without Following the Terms of the Applicable Licenses

48.   Below is an explanation of how Codex functions. When Codex is prompted with:

```
function isEven(n) {
```

it assumes this is the beginning of a function written in the JavaScript language that will test
whether a number is even.

49.   Based on this assumption, Codex will then provide Output meant to complete the rest of the function. Based on the given prompt, it produced the following response:[7]

```javascript
function isEven(n) {
  if (n == 0)
    return true;
  else if (n == 1)
    return false;
  else if (n < 0)
    return isEven(-n);
  else
    return isEven(n - 2);
}
console.log(isEven(50));
// → true
console.log(isEven(75));
// → false
console.log(isEven(-1));
// → ??
```

50.   The function itself occupies the first ten lines. Six additional lines follow the function, beginning with "`console.log(isEven(50))`". On possible explanation for Codex's inclusion of these lines is to test the "`isEven`" function. Though not part of the function itself, the lines will confirm the function works for certain values. In this case, the code implies that "`isEven(50)`" should return the value "`true`", and "`isEven(75)`" should return "`false`". Those answers are correct.

---

[7] Due to the nature of Codex, Copilot, and AI in general, Plaintiffs cannot be certain these examples would produce the same results if attempted following additional trainings of Codex and/or Copilot. However, these examples are representative of Codex and Copilot's Output at the time just prior to the filing of this Complaint.

51. The penultimate line indicates "`isEven(-1)`" should return "`??`". This is an error, as "`isEven(-1)`" should return "`false`".

52. Codex cannot and does not understand the meaning of software code or any other Licensed Materials. But in training, what became Codex was exposed to an enormous amount of existing software code (its "Training Data") and—with input from its trainers and its own internal processes—inferred certain statistical patterns governing the structure of code and other Licensed Materials. The finished version of Codex, once trained, is known as a "Model."

53. When given a prompt, such as the initial prompt discussed above—"`function isEven(n) {`"—Codex identifies the most statistically likely completion, based on the examples it reviewed in training. Every instance of Output from Codex is derived from material in its Training Data. Most of its Training Data consisted of Licensed Materials.

54. Codex does not "write" code the way a human would, because it does not understand the meaning of code. Codex's lack of understanding of code is evidenced when it emits extra code that is not relevant under the circumstances. Here, Codex was only prompted to produce a function called "`isEven`". To produce its answer, Codex relied on Training Data that also appended the extra testing lines. Having encountered this function and the follow-up lines together frequently, Codex extrapolates they are all part of one function. A human with even a basic understanding of how JavaScript works would know the extra lines aren't part of the function itself.

55. Beyond the superfluous and inaccurate extra lines, this "`isEven`" function also contains two major defects. First, it assumes the variable "n" holds an integer. It could contain some other kind of value, like a decimal number or text string, which would cause an error. Second, even if "n" does hold an integer, the function will trigger a memory error called a "stack overflow" for sufficiently large integers. For these reasons, experienced programmers would not use Codex's Output.

56. Codex does not identify the owner of the copyright to this Output, nor any other—it has not been trained to provide Attribution. Nor does it include a Copyright Notice nor any License Terms attached to the Output. This is by design—Codex was not coded or trained to

1   track or reproduce such data. The Output in the example above is taken from *Eloquent Javascript*

2   by Marijn Haverbeke.[8]

3   57.   Here is the exercise from *Eloquent Javascript*:

4
```
// Your code here.
```

5

6
```
console.log(isEven(50));
```

7
```
// → true
```

8
```
console.log(isEven(75));
```

9
```
// → false
```

10
```
console.log(isEven(-1));
```

11
```
// → ??
```

12   58.   The exercise includes the "??" error. However, for Haverbeke's purposes, this is

13   not an error but a placeholder value for the reader to fill in. Codex—as a mere probabilistic

14   model—fails to recognize this nuance. The inclusion of the double question marks confirms

15   unequivocally that Codex took this code directly from a copyrighted source without following any

16   of the attendant License Terms.

17   59.   Haverbeke provides the following solution to the function discussed above:

18
```
function isEven(n) {
```

19
```
  if (n == 0) return true;
```

20
```
  else if (n == 1) return false;
```

21
```
  else if (n < 0) return isEven(-n);
```

22
```
  else return isEven(n - 2);
```

23
```
}
```

24

25
```
console.log(isEven(50));
```

26

[8] https://eloquentjavascript.net/code/#3.2. *Eloquent Javascript* is "Licensed under a Creative

27   Commons [A]ttribution-[N]oncommercial license. All code in this book may also be considered
licensed under an MIT license." *See* https://eloquentjavascript.net/. Thus, having also been

28   posted on GitHub, the code Codex relied on meets the definition of Licensed Materials.

```
// → true
console.log(isEven(75));
// → false
console.log(isEven(-1));
// → false
```

60.   Aside from different line breaks—which are not semantically meaningful in JavaScript—this code for the function "`isEven`" is the same as what Codex produced. The tests are also the same, though in this case Haverbeke provides the right answer for "`isEven(-1)`", which is "`false`". Codex has reproduced Haverbeke's Licensed Material almost verbatim, with the only difference being drawn from a different portion of those same Licensed Materials.

61.   There are many copies of Haverbeke's code stored in public repositories on GitHub, where programmers who are working through Haverbeke's book store their answers.

62.   The MIT license provides that "The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software."[9] Any person taking this code directly from *Eloquent JavaScript* would have direct access to these License Terms and know to follow them if incorporating the Licensed Materials into a derivative work and/or copying them. Codex does not provide these License Terms.

63.   OpenAI Codex's Output would frequently, perhaps even constantly, contain Licensed Materials, i.e., it would have conditions associated with it through its associated license. In its 2021 research paper about Codex called "Evaluating Large Language Models Trained on Code," OpenAI stated Codex's Output is "often incorrect" and can contain security vulnerabilities and other "misalignments" (meaning, departures from what the user requested).

64.   Most open-source licenses require attribution of the author, notice of their copyright, and a copy of the license specifically to ensure that future coders can easily credit all previous authors and ensure they adhere to all applicable licenses. All the Suggested Licenses include these requirements.

---

[9] *See* Appendix A for full text of the MIT License.

65.   Ultimately, Codex derives its value primarily from its ability to locate and output potentially useful Licensed Materials. And from its obfuscation of any rights associated with those materials.

C.   Copilot Outputs Copyrighted Materials Without Following the Terms of the Applicable Licenses

66.   GitHub Copilot works in a similar way to OpenAI Codex. As mentioned above, a modified version of Codex is used as the engine that powers Copilot.

67.   Copilot is installed by the end user as an extension to various code editors, including Microsoft's Visual Studio and VS Code. As the user types into the editor, their code is uploaded in real time to Microsoft's Azure cloud platform, where they become prompts for Copilot.

68.   When we give Copilot the same prompt discussed above in Paragraph 48, "`function isEven(n) {`", it interprets the prompt as the beginning of a function written in the JavaScript language that will test whether a number is even, same as Codex.

69.   However, the Output of Copilot in response to the prompt is different than Codex, namely:

```
function isEven(n) {
    return n % 2 === 0;
}
```

70.   This function is much closer to what a human programmer might use as compared to Codex's older, inaccurate offering. It handles all values and types of "n" correctly. It does not cause a stack overflow for larger values of "n" like the Codex Output.

71.   Copilot's Output, like Codex's, is derived from existing code. Namely, sample code that appears in the online book *Mastering JS*, written by Valeri Karpov.[10] Like Codex's Output, Copilot's is also based upon copyrighted educational material. *Mastering JS* is a set of educational exercises for programmers. Like *Eloquent Javascript*, there are many copies of

---

[10] https://masteringjs.io/tutorials/fundamentals/modulus/.

1 Karpov's exercise stored in public repositories on GitHub. Programmers working through

2 *Mastering JS* store their answers there.

3   72.   If Copilot is prompted with the name of a function that will test whether a number

4 is prime (that is, a number that can only be evenly divided by 1 and itself), namely "`function`

5 `isPrime(n) {`", it returns:

```
function isPrime(n) {
    if (n < 2) {
        return false;
    }
    for (let i = 2; i < n; i++) {
        if (n % i === 0) {
            return false;
        }
    }
    return true;
}
```

17   73.   Though this function will work, it contains an error often made by beginner

18 programmers that makes it much slower than it could be. Namely, the loop in the middle, which

19 checks possible divisors, does not need to check every divisor smaller than "n," only the divisors

20 smaller than the square root of "n". As with Codex, Copilot has no understanding of how the

21 code works. It knows that more functions called "`isPrime`" contain the portion that checks for

22 all divisors smaller than "n", so that is what it offers. It does not return what it "thinks" is best, it

23 returns what it has seen *the most*. It is not writing, it is reproducing (i.e., copying).

24   74.   Like the other examples above—and most of Copilot's Output—this output is

25 nearly a verbatim copy of copyrighted code. In this case, it is substantially similar to the

26 "`isPrime`" function in the book *Think JavaScript* by Matthew X. Curinga et al,[11] which is:

27

28 [11] https://matt.curinga.com/think-js/#solving-problems-with-for-loops.

```
function isPrime(n) {
    if (n < 2) {
        return false;
    }
    for (let i = 2; i < n; i++) {
        if (n % i === 0) {
            return false;
        }
    }
    return true;
}
```

75.   As with the other examples above, the source of Copilot's Output is a programming textbook. Also like the books the other examples were taken from, there are many copies of Curinga's code stored in public repositories on GitHub where programmers who are working through Curinga's book keep copies of their answers.

76.   The material in Curinga's book is made available under the GNU Free Documentation License. Although this is not one of the Suggested Licenses, it contains similar attribution provisions, namely that "You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License."[12]

77.   As with Codex, Copilot does not provide the end user any attribution of the original author of the code, nor anything about their license requirements. There is no way for the Copilot user to know that they must provide attribution, copyright notice, nor a copy of the license's text. And with regard to the GNU Free Documentation License, Copilot users would not be aware that they are limited in what conditions they can place on the use of derivative works

---

[12] https://matt.curinga.com/think-js/#gnu-free-documentation-license.

they make using this copyrighted code. Had the Copilot user found this code in a public GitHub

repository or a copy of the book it was originally published in, they would find the GNU Free

Documentation License at the same time and be aware of its terms. Copilot finds that code for the

user but excises the license terms, copyright notice, and attribution. This practice allows its users

to assume that the code can be used without restriction. It cannot.

D.   Codex and Copilot Were Trained on Copyrighted Materials Offered Under Licenses

78.   Codex is an AI system. Another way to describe it is a "model." Without Codex,

Copilot, or another AI-code-lookup-tool, code is written both by originating code from the

writer's own knowledge of how to write code as well as by finding pre-written portions of code

that—under the terms of the applicable license—may be incorporated into the coding project.

79.   Unlike a human programmer that has learned how code works and notices when

code it is copying has attached license terms, a copyright notice, and/or attribution, Codex and

Copilot were developed by feeding a corpus of material, called "training data," into them. These

AI programs ingest all the data and, through a complex probabilistic process, predict what the

most likely solution to a given prompt a user would input is. Though more complicated in

practice, essentially Copilot returns the solution it has found in the most projects when those

projects are somehow weighted to adjust for whatever variables Codex or Copilot have identified

as relevant.

80.   Codex and Copilot were not programmed to treat attribution, copyright notices,

and license terms as legally essential. Defendants made a deliberate choice to expedite the release

of Copilot rather than ensure it would not provide unlawful Output.

81.   The words "study" and "training" and "learning" in connection with AI describe

algorithmic processes that are not analogous to human reasoning. An AI models cannot "learn"

as humans do, nor can it "understand" semantics and context the way humans do. Rather, it

detects statistically significant patterns in its training data and provides Output derived from its

training data when statistically appropriate. A "brute force" approach like this would not be

efficient nor even possible for humans. A human could not memorize, statistically analyze, and

easily access thousands of gigabytes of existing code, a task now possible for powerful computers

like those that make up Microsoft's Azure cloud platform. To accomplish the same task, a human

may search for Licensed Materials that serve their purpose if they believe such materials exist.

And if that human finds such materials, they will probably abide by its License Terms rather than

risk infringing its owners' rights. At the very least, if they incorporate those Licensed Materials

into their own project without following its terms they will be doing so knowingly.

E.   Copilot Was Launched Despite Its Propensity for Producing Unlawful Outputs

82.   GitHub and OpenAI have not provided much detail regarding what data Codex

and OpenAI were trained on. Plaintiffs know for certain from GitHub and OpenAI's statements,

that both systems were trained on publicly available GitHub repositories, with Copilot having

been trained on all available public GitHub repositories. Thus, if Licensed Materials have been

posted to a GitHub public repository, Plaintiffs and the Class can be reasonably certain it was

ingested by Copilot and is sometimes returned to users as Output.

83.   According to OpenAI, Codex was trained on "billions of lines of source code from

publicly available sources, including code in public GitHub repositories". Similarly, GitHub has

described[13] Copilot's training material as "billions of lines of public code." GitHub researcher

Eddie Aftandilian confirmed in a recent podcast[14] that Copilot is "train[ed] on public repos on

GitHub."

84.   In a recent customer-support message, GitHub's support department clarified

certain facts about training Copilot. First, GitHub said that "training for Codex (the model used

by Copilot) is done by OpenAI, not GitHub." Second, in its support message, GitHub put

forward a more detailed justification for its use of copyrighted code as training data:

---

[13] https://github.blog/2021-06-30-github-copilot-research-recitation/.

[14] https://www.se-radio.net/2022/10/episode-533-eddie-aftandilian-on-github-copilot/.

> Training machine learning models on publicly available data is considered fair use across the machine learning community . . . OpenAI's training of Codex is done in accordance with global copyright laws which permit the use of publicly accessible materials for computational analysis and training of machine learning models, and do not require consent of the owner of such materials. Such laws are intended to benefit society by enabling machines to learn and understand using copyrighted works, much as humans have done throughout history, and to ensure public benefit, these rights cannot generally be restricted by owners who have chosen to make their materials publicly accessible.

The claim that training ML models on publicly available code is widely accepted as fair use is not true. And regardless of this concept's level of acceptance in "the machine learning community," under Federal law, it is illegal.

85.   Former GitHub CEO Nat Friedman said in June 2021—when Copilot was released to a limited number of customers—that "training ML systems on public data is fair use."[15] Friedman's statement is pure speculation; no Court has considered the question of whether "training ML systems on public data is fair use." The Fair Use affirmative defense is only applicable to Section 501 copyright infringement. It is not a defense to violations of the DMCA, Breach of Contract, nor any other claim alleged herein. It cannot be used to avoid liability here. At the same time Friedman asserted "the output [of Copilot] belongs to the operator."

86.   Other open-source stakeholders have made this point already. For example, in June 2021, Software Freedom Conservancy ("SFC"), a prominent open-source advocacy organization, asked Microsoft and GitHub to provide "legal references for GitHub's public legal positions." No references were provided by any of the Defendants.[16]

87.   Beyond the examples above, Copilot regularly Output's verbatim copies of Licensed Materials. For example, Copilot reproduced verbatim well-known code from the game Quake III, use of which is governed by one of the Suggested Licenses—GPL-2.[17]

---

[15] https://twitter.com/natfriedman/status/1409914420579344385/.

[16] https://sfconservancy.org/blog/2022/feb/03/github-copilot-copyleft-gpl/.

[17] https://twitter.com/stefankarpinski/status/1410971061181681674/.

88.   Copilot also reproduced code that had been released under a license that allowed its use only for free games and required attribution by including a copy of the license. Copilot did not mention nor include the underlying license when providing a copy of this code as Output.[18]

89.   Texas A&M computer-science professor Tim Davis has provided numerous examples of Copilot reproducing code belonging to him without its license or attribution.[19]

90.   GitHub concedes that in ordinary use, Copilot will reproduce passages of code verbatim: "Our latest internal research shows that about 1% of the time, a suggestion [Output] may contain some code snippets longer than ~150 characters that matches" code from the training data. This standard is more limited than is necessary for copyright infringement. But even using GitHub's own metric and the most conservative possible criteria, Copilot has violated the DMCA at least tens of thousands of times.

91.   In June 2022, Copilot had 1,200,000 users. If only 1% of users have ever received Output based on Licensed Materials and only once each, Defendants have "only" breached Plaintiffs' and the Class's Licenses 12,000 times. However, each time Copilot outputs Licensed Materials without attribution, the copyright notice, or the License Terms it violates the DMCA three times. Thus, even using this extreme underestimate, Copilot has "only" violated the DMCA 36,000 times.[20] Because Copilot constantly Outputs code as a user writes, and because nearly all of Copilot's training data was Licensed Material, this number is most likely exponentially lower than the true number of breaches and DMCA violations.

F.   Open-Source Licenses Began to Appear in the Early 1990s

92.   In 1991, software engineer Linus Torvalds began a project to create a UNIX-like operating system that would run on common PC hardware. This project became known as Linux.

---

[18] https://twitter.com/ChrisGr93091552/status/1539731632931803137/.

[19] https://twitter.com/DocSparse/status/1581461734665367554/.

[20] These violations of Section 1202 of the DMCA each incur statutory damages of "not less than $2,500 or more than $25,000." 17 U.S.C. § 1203(c)(3)(B). This extremely conservative estimate of Defendants' number of direct violations translates to $90 million to $900 million in statutory damages.

93.   To encourage adoption of his system, and persuade other programmers to contribute, he released Linux under what was then an unusual software license called the GNU General Public License, or GPL.

94.   The GPL is a software license. But whereas most software licenses required payment, software under the GPL is provided for free. Whereas most software licenses did not include source code, GPL software always included source code. And whereas most software licenses prohibited derivative works, the GPL not only allowed it, but encouraged it.

95.   In certain ways, however, the GPL still operated like a traditional software license. For example, consistent with copyright law, it depended on an assertion of copyright by the software author. Even though GPL software was available at no charge, the GPL contained conditions on its users as licensees.

96.   One license requirement was that a program derived from GPL software had to redistribute certain information about that software:

> You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this General Public License and to the absence of any warranty; and give any other recipients of the Program a copy of this General Public License along with the Program.[21]

Failure to adhere to these conditions constituted a violation of the license, triggering the possibility of legal action. Provisions of the GPL are enforceable, and many GPL licensors have sought to enforce GPL licenses though court proceedings and other litigation.

97.   The early years of Linux paralleled the early years of the World Wide Web. The fact that Linux was free and ran on common computer hardware made it a popular choice for web servers. Because of its contrarian GPL licensing, Linux became hugely popular. A large ecosystem of other programs and tools grew around it. This contributed to the explosive growth of the web and other network services across the rest of the 1990s.

---

[21] https://www.gnu.org/licenses/old-licenses/gpl-1.0.en.html.

98.   In turn, the growth of the World Wide Web made it easier for developers in different places to collaborate on software. The GPL, and licenses like it, were a natural fit for this kind of collaborative work.

99.   Around 1998, a new name was coined as an umbrella term for these principles of software licensing and development: *open source*.

G.   Microsoft Has a History of Flouting Open-Source License Requirements

100.   During the 1980s and 1990s, Microsoft was primarily a software company, focusing largely on operating systems and related applications. These included its DOS operating system and later, its Windows operating system. Windows generated billions of dollars in revenue from its sale and licensing as proprietary software for desktop computers and servers. Microsoft derived substantial income from sale of licensed products and devotes substantial resources to protecting and enforcing such licenses.

101.   Windows is a graphical operating system. It allows users to view and store files, run software and games, play videos, and provides a way to connect to the internet.

102.   Linux represented a competitive threat to Windows. It ran on the same hardware. It performed many of the same functions. It was free. Many programmers at the time considered Linux to be functionally superior to Windows.

103.   Microsoft has engaged in a problematic practice known as "vaporware," where products are announced but are in fact late, never manufactured, or canceled. Typically the company promising vaporware never has any intention of providing it. The term vaporware was coined by Microsoft in 1982 in reference to the development of its Xenix operating system.

104.   Microsoft described its anti-Linux strategy as "FUD," standing for fear, uncertainty, and doubt. Microsoft focused extra attention to Linux's open-source aspects.

105.   In 1998, a source at Microsoft leaked what became known as the "Halloween Documents", revealing Microsoft's thinking on how to counter the competitive threat from Linux. Among other things, the documents emphasized the importance of countering the "long

term developer mindshare threat", and concluded that to defeat open source, "[Microsoft] must target a process rather than a company".[22]

106.   In 2001, Microsoft CEO Steve Ballmer said "The way the [GPL] is written, if you use any open-source software, you must make the rest of your software open source. . . . Linux is a cancer that attaches itself in an intellectual property sense to everything it touches."[23] Ballmer's summary of GPL licensing was not accurate. In 2001, Linux was being used by corporations of every size. The growth of open source up to that point, and since, has been made possible by the open-source community's respect for and compliance with applicable licenses.

107.   In 2001, Microsoft was the defendant in a major software-related antitrust case, *United States v. Microsoft Corporation*.[24] In this case, the U.S. Department of Justice accused Microsoft of maintaining a software monopoly by illegally imposing technical restrictions on manufacturers of personal computers, including "tying" violations related to the Internet Explorer web browser. Judge Thomas Penfield Jackson, who presided over the antitrust trial, opined that Microsoft is "a company with an institutional disdain for both the truth and for rules of law that lesser entities must respect. It is also a company whose 'senior management' is not averse to offering specious testimony to support spurious defenses to claims of its wrongdoing."[25]

108.   In 2007, Microsoft admitted that it tried to influence the vote of an ISO open-standards committee by offering money to certain business partners in Sweden to vote for Microsoft's preferred outcome.[26]

109.   After observing the rapid growth of Amazon's original cloud computing products, Microsoft has expanded its business into cloud computing, which it has branded Microsoft Azure or simply Azure. Microsoft announced Azure to developers in 2008. It was formally released in

---

[22] http://www.catb.org/esr/halloween/halloween1.html.

[23] https://lwn.net/2001/0607/a/esr-big-lie.php3.

[24] No. Civ.A. 00–1457 TPJ.

[25] *Jackson v. Microsoft Corp.*, 135 F. Supp. 2d 38 (D.D.C. 2001).

[26] https://learn.microsoft.com/en-us/archive/blogs/jasonmatusow/open-xml-the-vote-in-sweden/.

1 2010. Azure uses large-scale virtualization at Microsoft data centers and offers many hundreds of

2 services, including infrastructure as a service ("IaaS"), platform as a service ("PaaS"), compute

3 services, Azure Active Directory, mobile services, storage services, communication services, data

4 management, messaging, developer services, Azure AI, blockchain, and others.

5 **H.   GitHub Was Designed to Cater to Open-Source Projects**

6   110.   By 2002, Linux had become immensely popular. But the project itself had become

7 unwieldy and had outgrown its reliance on informal systems of managing software source code

8 (also known as *source-control systems*). The Linux community needed something better.

9   111.   Linus Torvalds set about writing a new source-control system. He named his new

10 system Git. He released it under the GPL. It quickly became the source-control system of choice

11 for open-source programmers.

12   112.   A single software project stored in Git is called a *source repository*, commonly

13 shortened to *repository* or just *repo*. A Git source repository would typically be stored on a

14 networked server accessible to a group of programmers.

15   113.   This became less convenient, however, when programmers were distributed

16 among multiple locations, rather than being in a single location. A Git repository could be stored

17 on an internet-accessible server. But setting up that server hardware and being responsible for it

18 was inconvenient and expensive.

19   114.   In 2008, a group of open-source developers in San Francisco, California founded

20 GitHub. GitHub managed internet servers that hosted Git source repositories. With an account at

21 GitHub, an open-source developer could easily set up a Git project accessible to collaborators

22 anywhere in the world. From early on, GitHub's core market was open-source developers, whom

23 it attracted by making many of its hosting services free.

24   115.   Most open-source programmers used GitHub to create "public" repositories,

25 meaning that anyone could view them & access them. GitHub also allowed programmers and

26 organizations to create "private" repositories, which were not accessible from the public GitHub

27 website, and required password access.

28

116. Open-source licensing was integral to GitHub. GitHub encouraged open-source developers to understand and use open-source licenses for their work. Many—though not all—public repositories on GitHub carry an open-source license. By convention, this license is stored at the top level of each repository in a file called `LICENSE`. GitHub's interface also includes a button on the front pages of most repositories users can click to see details of the applicable license. A human user could easily find the license in either of these locations—as could an AI anywhere near as powerful as Codex or Copilot.

117. Though the GPL is one of the early open-source licenses and remains common, it's not the only open-source license. Examples of other common open-source licenses include the MIT License, the Apache License, and the Berkeley Software Distribution License (all of which are included in the Suggested Licenses).

118. Though these licenses differ in their wording and their details, most of them share a requirement that a copy of the license be included with any copy, derivative, or redistribution of the software, and that the author's name and copyright notice remains intact. This is not a controversial requirement of open-source licenses—indeed, it has been an integral part of the GPL for over 30 years.

119. There are also many public repositories on GitHub that have no license. Though GitHub has encouraged awareness of licenses among its users, it has never imposed a default license on public repositories. A public repository without a license is subject to ordinary rules of U.S. copyright.

120. Open-source developers flocked to GitHub. By 2018, GitHub had become the largest and most successful Git hosting service, hosting millions of users and projects.

121. In October 2018, Microsoft acquired GitHub for $7.5 billion. It was important to Microsoft that programmers use GitHub. Microsoft had developed a well-deserved poor reputation because of its documented vaporware, FUD, and other business practices, including those targeted at open-source programs and programming, and open-source licensing specifically. Microsoft made false and misleading statements and omissions to assuage such concerns,

including its primary mantra intended to win over the open-source community: "Microsoft Loves Open Source."

I.   OpenAI Is Intertwined with Microsoft and GitHub

122.   OpenAI, Inc. is a nonprofit corporation founded in December 2015 by a group that included Greg Brockman, Ilya Sutskever, and other AI researchers; Elon Musk, CEO of Tesla; and Sam Altman, president of Y Combinator, a tech-startup incubator with hundreds of companies in its portfolio. Musk and Altman served as co-chairs of OpenAI, Inc. One of OpenAI, Inc.'s current board members is Reid Hoffman, founder of LinkedIn, which is now a Microsoft subsidiary. Mr. Hoffman is also a member of the Microsoft Board of Directors.

123.   Less than a year later, in November 2016, it first partnered with Microsoft. It described the partnership as follows: "We're working with Microsoft to start running most of our large-scale experiments on Azure. This will make Azure the primary cloud platform that OpenAI is using for deep learning and AI, and will let us conduct more research and share the results with the world."

124.   Initially, OpenAI, Inc. held itself out as a "non-profit artificial intelligence research company" that sought to shape AI "in the way that is most likely to benefit humanity as a whole."

125.   OpenAI, Inc. reportedly secured $1 billion in initial funding, from sources that were largely not disclosed, but included at least most of its founders.

126.   OpenAI, Inc. obtained its initial source of training data from its founders' companies. According to reporting at the time, Musk and Altman planned to "pool[] online data from their respective companies" to serve as training data for OpenAI, Inc. projects. Musk planned to contribute data from Tesla; Altman planned to have Y Combinator companies "share their data with OpenAI."[27]

127.   In February 2019, Altman created OpenAI, LP, a for-profit subsidiary of the nonprofit entity OpenAI, Inc. The new OpenAI, LP entity would serve as a vessel for accepting traditional outside investment in exchange for equity and distributing profits.

---

[27] https://www.wired.com/2015/12/elon-musks-billion-dollar-ai-plan-is-about-far-more-than-saving-the-world/.

1   128.   In July 2019, OpenAI, L.P. accepted a $1 billion investment from Microsoft. In

2   addition to cash, Microsoft would become the exclusive licensor of certain OpenAI, LP products

3   (including GPT-3, described below in Paragraph 131). Also, as part of this alliance, OpenAI, LP

4   would use Microsoft's cloud-computing platform, Azure, exclusively to develop and host its

5   products. Some portion of Microsoft's investment was paid in credits for use of Azure rather

6   than cash. Finally, Microsoft and OpenAI agreed to "jointly build new Azure AI supercomputing

7   technologies."

8   129.   Azure is a major growth area for Microsoft. In its most recent earnings report on

9   October 25, 2022, "Azure and other cloud services" grew by 35% from the previous quarter, more

10 than any other product.[28] Azure has grown rapidly since Microsoft began its partnership with

11 OpenAI in 2016. Its revenue grew by 50% or more every quarter from 2016 through the first three

12 quarters of 2020.

13 130.   In May 2020, Microsoft and OpenAI announced they had jointly built a

14 supercomputer in Azure that would be used exclusively by OpenAI to train its AI models.

15 Microsoft's influence over and frequent collaboration with OpenAI has led some to describe

16 Microsoft as "the unofficial owner of OpenAI."[29]

17 131.   One of OpenAI's projects is GPT-3, a so-called "large language model" designed

18 to emit naturalistic text. When researchers noticed that GPT-3 could also generate software code,

19 they started studying whether they could make a new AI model specifically trained for this

20 purpose. This project became known as Codex.

21 132.   Sometime after July 2019, OpenAI and Microsoft began collaborating on a code-

22 completion product for GitHub that would use Codex as its underlying model. This product

23 became known as Copilot.

24 133.   On September 28, 2022, OpenAI released an image-generation AI called DALL-

25 E-2. Much like Copilot, DALL-E-2 removes any attribution and/or copyright notice from the

26

27 [28] https://www.microsoft.com/en-us/Investor/earnings/FY-2023-Q1/press-release-webcast/.

28 [29] https://venturebeat.com/ai/what-to-expect-from-openais-codex-api/.

1   images it uses to create derivative works. Like with Codex, here, OpenAI ignores the rights of the

2   owners of copyrights to images it has ingested.

3   134.   In another joint project, Microsoft and OpenAI recently launched a preview of a

4   product called "Azure OpenAI Service."[30] This service will "Leverage large-scale, generative AI

5   models with deep understandings of language and code to enable new reasoning and

6   comprehension capabilities for building cutting-edge applications. Apply these coding and

7   language models to a variety of use cases, such as writing assistance, code generation, and

8   reasoning over data. Detect and mitigate harmful use with built-in responsible AI and access

9   enterprise-grade Azure security."

10 **J.   Conclusion of Factual Allegations**

11   135.   Future AI products may represent a bold and innovative step forward. GitHub

12 Copilot and OpenAI Codex, however, do not. Defendants should not have released these

13 products until they could ensure that they did not constantly violate Plaintiffs' and the Class's

14 intellectual-property rights, licenses, and other rights.

15   136.   Defendants have made no attempt to comply with the open-source licenses that

16 are attached to much of their training data. Instead, they have pretended those licenses do not

17 exist, and trained Codex and Copilot to do the same. By simultaneously violating the open-source

18 licenses of tens-of-thousands—possibly millions—of software developers, Defendants have

19 accomplished software piracy on an unprecedented scale. As Microsoft's Co-Founder Bill Gates

20 once said regarding software piracy: "the thing you do is theft."[31]

21   137.   There is no inherent limitation or constraint of AI systems that made any of this

22 necessary. Defendants chose to build AI systems designed to enhance their own profit at the

23 expense of a global open-source community that they had once sought to foster and protect.

24 GitHub and OpenAI are profiting at the expense of Plaintiffs' and the Class's rights.

25

26

---

27 [30] https://azure.microsoft.com/en-us/products/cognitive-services/openai-service/.

28 [31] https://www.digibarn.com/collections/newsletters/homebrew/V2_01/gatesletter.html

1

VIII. CLAIMS FOR RELIEF

2

COUNT I
VIOLATION OF THE DIGITAL MILLENIUM COPYRIGHT ACT
3
17 U.S.C. §§ 1201–1205
(Direct, Vicarious, and Contributory)
4
(Against All Defendants)

5   138.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding

6 and succeeding paragraph as though fully set forth herein.

7   139.   Plaintiffs and members of the Class own the copyrights to Licensed Materials used

8 to train Codex and Copilot. Copilot was trained on millions—possibly billions—of lines of code

9 publicly available on GitHub. Copilot runs on Microsoft's Azure cloud platform exclusively and

10 Microsoft had input in the creation of Copilot. Microsoft is aware that Copilot ignores License

11 Terms and that it was trained almost exclusively on Licensed Materials.

12   140.   Plaintiffs and members of the Class included the following Copyright

13 Management Information (as defined in Section 1202(c) of the DMCA) ("CMI") in the Licensed

14 Materials:

15   a.   copyright notices;

16   b.   the title and other information identifying the Licensed Materials;

17   c.   the name of, and other identifying information about, the authors of the Licensed

18   Materials;

19   d.   the name of, and other identifying information about, the copyright owners of the

20   Licensed Materials;

21   e.   terms and conditions for use of the Licensed Materials, specifically the Suggested

22   Licenses; and

23   f.   identifying numbers or symbols referring to CMI or links to CMI.

24   141.   Defendants did not contact Plaintiffs and the Class to obtain authority to remove

25 or alter CMI from the Licensed Materials within the meaning of the DMCA.

26   142.   Defendants knew that they did not contact Plaintiffs and the Class to obtain

27 authority to remove or alter CMI from the Licensed Materials within the meaning of the DMCA.

28

143.   As part of the scheme, Defendants did not attempt to contact Plaintiffs to obtain authority to remove or alter CMI from the Licensed Materials within the meaning of the DMCA. In fact, Defendants' removal of CMI made it difficult or impossible to contact Plaintiffs and the Class to obtain authority to remove or alter CMI from the Licensed Materials within the meaning of the DMCA. Rather, Defendants removed or altered CMI from open-source code that is owned by Plaintiffs and the Class after the code was uploaded to a GitHub repository by incorporating it into Copilot with its CMI removed.

144.   Without the authority of Plaintiffs and the Class, Defendants intentionally removed or altered CMI from the Licensed Materials after they were uploaded to one or more GitHub repositories.

145.   Defendants had access to but were not licensed by Plaintiffs nor the Class to train any machine learning, AI, or other pseudo-intelligent computer program, algorithm, or other functional prediction engine using the Licensed Materials.

146.   Defendants had access to but were not licensed by Plaintiffs nor the Class to incorporate the Licensed Materials into Copilot.

147.   Defendants had access to but were not licensed by Plaintiffs nor the Class to create Derivative Works[32] based upon the Licensed Materials.

148.   Defendants had access to but were not licensed by Plaintiffs nor the Class to distribute the Licensed Materials as they do through Copilot.

149.   Without the authority of Plaintiffs and the Class, Defendants distributed CMI knowing that the CMI had been removed or altered without authority of the copyright owner or the law with respect to the Licensed Materials.

150.   Defendants distributed copies of the Licensed Materials knowing and intending that CMI had been removed or altered without authority of the copyright owner or the law, with respect to the Licensed Materials.

---

[32] "Derivative Works" as used herein refers to Copilot's Output to the extent they are derived from Licensed Materials. The definition also includes the Copilot product itself, which is a Derivative Work based upon a large corpus of Licensed Materials.

1   151.   Defendants removed or altered CMI from the Licensed Materials knowing and

2   intending that it would induce, enable, facilitate, or conceal infringement of copyright.

3   152.   Without the CMI associated with the Licensed Materials, Copilot users are

4   induced or enabled to copy the Licensed Materials. Because CMI has been removed, Copilot

5   users do not know whether Output is owned by someone else and subject to restrictions on use.

6   Without the CMI, copyright infringement is facilitated or concealed, because Plaintiffs and the

7   Class are prevented from knowing or learning that the Output is based upon one or more of the

8   Licensed Materials. Use of the Licensed Materials is not infringement when the terms of the

9   applicable Suggested License are followed. Had the CMI not been removed, Copilot users would

10   be aware of the Licenses and their obligations under them. The terms of the applicable Suggested

11   License would have allowed those users to use the Licensed Materials without infringement. By

12   withholding and concealing license information and other CMI, Defendants prevented Copilot

13   users from making non-infringing use of the Licensed Materials. This contradicts the express

14   wishes of Plaintiffs and the Class, which are set forth explicitly in the Suggested Licenses under

15   which the Licensed Materials are offered.

16   153.   Defendants removed or altered CMI from Licensed Materials owned by Plaintiffs

17   and the Class while possessing reasonable grounds to know that it would induce, enable, facilitate,

18   and/or conceal infringement of copyright in violation of the DMCA. By omitting and concealing

19   CMI from Copilot's Output, Defendants have reasonable grounds to know that innocent

20   infringers are induced or enabled to copy the Licensed Materials, because CMI has been

21   removed. Without the CMI, Defendants have reasonable grounds to know copyright infringement

22   is facilitated or concealed, because Plaintiffs and the Class have the difficult or impossible task of

23   proving the Licensed Materials belong to them.

24   154.   Defendants knowingly provided CMI that is false with respect to the Licensed

25   Materials. Defendants have a business practice of asserting and/or implying that Copilot is the

26   author of the Licensed Materials. Defendants knowingly distributed CMI that is false, with

27   respect to the Licensed Materials. Defendants have a business practice of asserting and/or

28   implying that Copilot is the author of the Licensed Materials.

155.   Defendants provided or distributed false CMI from the Licensed Materials with respect to Copilot's Output with the intent and foreseeable result to induce, enable, facilitate, or conceal infringement. Defendants have a business practice of asserting and/or implying that Copilot is the author of the Licensed Materials. This false CMI induces or enables Defendants or Copilot users to copy the Licensed Materials. Defendants' false description of the source of Copilot's Output facilitated or concealed infringement by Defendants and Copilot users because Plaintiffs and the Class have the difficult or impossible task of proving that the copyrights to the suggested portions of their Licensed Materials belong to them once those Licensed Materials have been delinked from all identifying information and all license terms governing their use.

156.   The profits attributable to Defendants' violation of the DMCA include the revenue from: Copilot subscription fees, sales of or subscriptions to Defendants' Copilot-related products and/or services that are used to run Copilot, hosting Copilot on Azure, and any other of Defendants' products that contain copies of the Licensed Materials without all the original CMI. The Licensed Materials add nearly all value to the Copilot product because the purpose of Copilot is to provide code and the source of that code is the Licensed Materials. Without the Licensed Materials, Copilot would not be functional.

157.   On information and belief, Defendants could have trained Copilot to include attribution, copyright notices, and license terms when it provides Output covered by a License.

158.   Defendants did not request or obtain permission from Plaintiffs and the Class to use the Licensed Materials for Defendants' Copilot product.

159.   Defendants use of the Licensed Materials does not follow the requirements of the Suggested Licenses associated with the Licensed Materials. In particular, Copilot fails to provide attribution for the creator nor the owner of the Work. Copilot fails to include the required copyright notice included in the License. Copilot fails to include the applicable Suggested License's text.

160.   Defendants are sophisticated with respect to intellectual property matters related to open-source code. Microsoft in particular has extensive experience granting licenses, obtaining licenses, and enforcing license terms. Its most recent Annual Report states:

We protect our intellectual property investments in a variety of ways. We work actively in the U.S. and internationally to ensure the enforcement of copyright, trademark, trade secret, and other protections that apply to our software and hardware products, services, business plans, and branding. We are a leader among technology companies in pursuing patents and currently have a portfolio of over 69,000 U.S. and international patents issued and over 19,000 pending worldwide. While we employ much of our internally-developed intellectual property exclusively in our products and services, we also engage in outbound licensing of specific patented technologies that are incorporated into licensees' products. From time to time, we enter into broader cross-license agreements with other technology companies covering entire groups of patents. We may also purchase or license technology that we incorporate into our products and services. At times, we make select intellectual property broadly available at no or low cost to achieve a strategic objective, such as promoting industry standards, advancing interoperability, supporting societal and/or environmental efforts, or attracting and enabling our external development community. **Our increasing engagement with open source software will also cause us to license our intellectual property rights broadly in certain situations.**

Microsoft Corporation Annual Report, Form 10-K at 27 (July 28, 2022) (emphasis added).[33]

161. GitHub, which offers the Copilot product jointly with OpenAI, also has extensive experience with the DMCA. GitHub knows or reasonably should know that the Licensed Materials it hosts are subject to copyright. It provides the language of the Suggested Licenses to users, all of which include copyright notices. Its 2022 Transparency Report—January to June[34] states: "Copyright-related takedowns (which we often refer to as DMCA takedowns) are particularly relevant to GitHub because so much of our users' content is software code and can be eligible for copyright protection."[35] In the first six months of 2022, GitHub processed 1220 DMCA takedown requests. Its DMCA Takedown Policy[36] notes "GitHub probably never would have existed without the DMCA."

---

[33] https://microsoft.gcs-web.com/static-files/07cf3c30-cfc3-4567-b20f-f4b0f0bd5087/.

[34] https://github.blog/2022-08-16-2022-transparency-report-january-to-june/.

[35] https://github.blog/2022-08-16-2022-transparency-report-january-to-june/.

[36] https://docs.github.com/en/site-policy/content-removal-policies/dmca-takedown-policy#what-is-the-dmca/.

162.   GitHub also knows or reasonably should know the portions of the DMCA giving rise to Plaintiffs' claim. In its 2021 Transparency Report, "Before removing content based on alleged circumvention of copyright controls (under Section 1201 of the US DMCA or similar laws in other countries), we carefully review both the legal and technical claims, and we sponsor a Developer Defense Fund to provide developers with meaningful access to legal resources."[37]

163.   GitHub is aware that Copilot's removal of CMI is illegal. For example, it states that "publishing or sharing tools that enable circumvention are not [permitted]"[38] and "Distributing tools that enable circumvention is prohibited, even if their use by developers falls under the exemption [for security research]."[39] GitHub has also frequently published articles discussing the DMCA, its application, and the Copyright Office's guidance on its scope and exceptions.[40]

164.   Unless Defendants are enjoined from violating the DMCA, Plaintiffs and the Class will suffer great and irreparable harm by depriving them of the right to identify and control the reproduction and/or distribution of their copyrighted works, to have the terms of their open-source licenses followed, and to pursue copyright-infringement remedies. Defendants will not be damaged if they are required to comply with the DMCA. Plaintiffs and the Class members are therefore entitled to an injunction barring Defendants from violating the DMCA and impounding any device or product that is in the custody or control of Defendants and that the court has reasonable cause to believe was involved in a violation of the DMCA.

165.   Plaintiffs and the Class are further entitled to recover from Defendants the actual or statutory damages Plaintiffs and the Class sustained pursuant to 17 U.S.C. § 1203(c) and for Plaintiffs' and the Class's costs and attorneys' fees in enforcing the Licenses. Plaintiffs and the Class are also entitled to recover as restitution from Defendants for any unjust enrichment,

---

[37] https://github.blog/2022-01-27-2021-transparency-report/.

[38] https://github.blog/2020-11-19-take-action-dmca-anti-circumvention-and-developer-innovation/#what-dmca-exemptions-do-not-do/.

[39] https://github.blog/2021-11-23-copyright-office-expands-security-research-rights/.

[40] *See, e.g.*, Footnotes 34–39.

1   including gains, profits, and advantages that Defendants have obtained as a result of their breach

2   of the Licenses.

3   166.   Defendants conspired together and acted jointly and in concert pursuant to their

4   scheme to commit the acts that violated the DMCA alleged herein.

5   167.   Defendants induced Copilot users to unknowingly violate the DMCA by

6   withholding attribution, licensing, and other information as described herein.

7   **COUNT II**
BREACH OF CONTRACT—OPEN-SOURCE LICENSE VIOLATIONS

8   **Common Law**
(Against All Defendants)

9

10   168.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding

11   and succeeding paragraph as though fully set forth herein.

12   169.   Plaintiffs and the Class offer code under various Licenses, the most common of

13   which are set forth in Appendix A. Use of each of the Licensed Materials is allowed only pursuant

14   to the terms of the applicable Suggested License.

15   170.   Plaintiffs and the Class granted Defendants a license to copy, distribute, and/or

16   create Derivative Works under the Suggested Licenses. Each of the Suggested Licenses requires

17   at least (1) that attribution be given to the owner of the Licensed Materials used, (2) inclusion of a

18   copyright notice for the Licensed Materials used, and (3) inclusion of the terms of the applicable

19   Suggested License. When providing Output, Copilot does not comply with any of these terms.

20   171.   Defendants accepted the terms of Plaintiffs' and the Class's Licenses when it used

21   the licensed code to create Copilot and when it incorporated the licensed code into Copilot. They

22   have accepted and continue to accept the applicable Licenses every time Copilot Output's

23   Plaintiffs' or the Class's copyrighted code. As such, contracts have been formed between

24   Defendants on the one hand and Plaintiffs and the Class on the other.

25   172.   Plaintiffs and the Class have performed each of the conditions, covenants, and

26   obligations imposed on them by the terms of the License associated with their Licensed

27   Materials.

28

173. Plaintiffs and members of the Class hold the copyright in the contents of one or more code repositories that have been hosted on GitHub's platform.

174. Plaintiffs and the Class have appended one of the Suggested Licenses to each of the Licensed Materials.

175. Plaintiffs and the Class did not know about, authorize, approve, or license the Defendants' use of the Licensed Materials in the matter at issue in this Complaint before they were used by Defendants.

176. Defendants have substantially and materially breached the applicable Licenses by failing to provide the source code of Copilot nor a written offer to provide the source code upon the request of each licensee.

177. Defendants have substantially and materially breached the applicable Licenses by failing to provide attribution to the creator and/or owner of the Licensed Materials.

178. Defendants have substantially and materially breached the applicable Licenses by failing to include copyright notices when Copilot Outputs copyrighted OS code.

179. Defendants have substantially and materially breached the applicable Licenses by failing to identify the License applicable to the Work and/or including its text when Copilot Outputs code including a portion of a Work.

180. Plaintiffs and the Class have suffered monetary damages as a result of Defendants' conduct.

181. The conduct of Defendants is causing and, unless enjoined and restrained by this Court, will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully be compensated or measured in money.

182. As a direct and proximate result of these material breaches by Defendants, Plaintiffs and the Class are entitled to an injunction requiring Defendants to comply with all the terms of any License governing use of code that was used to train Copilot, otherwise incorporated into Copilot, and/or reproduced as Output by Copilot.

183. Plaintiffs and the Class are further entitled to recover from Defendants the damages Plaintiffs and the Class sustained—including consequential damages—for Plaintiffs' and

1   the Class's costs in enforcing their contractual rights. Plaintiffs and the Class are also entitled to

2   recover as restitution from Defendants for any unjust enrichment, including gains, profits, and

3   advantages that Defendants have obtained as a result of their breach of contract.

COUNT III
TORTIOUS INTERFERENCE IN A CONTRACTUAL RELATIONSHIP
Common Law
(Against All Defendants)

7   184.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding

8   and succeeding paragraph as though fully set forth herein.

9   185.   Defendants have wrongfully interfered with the business interests and

10   expectations of Plaintiffs and the Class by improperly using Copilot to create Derivative Works

11   that compete against OSC.

12   186.   At GitHub's upcoming yearly conference, GitHub Universe 2022, it will host a

13   presentation called "How to compete with open source—and win."

14   187.   Plaintiffs and the Class have suffered monetary, reputational, and other damages

15   as a result of Defendants' conduct.

16   188.   The harm was the actual, proximate, intentional, direct, and foreseeable

17   consequence of Defendant's conduct.

18   189.   The conduct of Defendants is causing and, unless enjoined and restrained by this

19   Court, will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully

20   be compensated or measured in money.

COUNT IV
FRAUD
Common Law
(Against GitHub)

24   190.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding

25   and succeeding paragraph as though fully set forth herein.

26   191.   GitHub made certain representations to Plaintiffs and the Class to induce them to

27   publicly post their code on GitHub. Specifically, in both its Terms of Service and its Privacy

28   Statement, GitHub promises not to sell Licensed Materials or anything else uploaded to or shared

1   with GitHub. It also promises not to distribute Licensed Materials outside GitHub. As explained

2   above, Copilot operates on an individual's computer as an extension to their editor as well as on

3   Microsoft's Azure cloud platform. Neither are part of GitHub. It Outputs in the user's editor,

4   which is not part of GitHub.

5   192.   Plaintiffs and the Class relied upon those representations in choosing to upload

6   Licensed Materials to GitHub. GitHub has long held itself out as the best place to host open-

7   source code repositories. It has courted the business of users it expects will include Licenses with

8   their code. It facilitates this by allowing users to easily select the name of a license, including the

9   Suggested Licenses, when creating a repository rather than finding the text of the license and

10   adding it themselves. GitHub provides the terms, it can hardly claim to be unaware of what they

11   are or what they mean. If it didn't understand the requirements of a given Suggested License, it

12   would not have provided it as an option to its users.

13   193.   GitHub failed to honor its representations in creating and operating Copilot. It

14   sells Plaintiffs' and the Class's Licensed Materials as part of Copilot. It also distributes them. It

15   does so without following any of the License Terms.

16   194.   As such, GitHub failed to honor its representations in operating Copilot.

17   195.   The conduct of GitHub is causing and, unless enjoined and restrained by this

18   Court, will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully

19   be compensated or measured in money. Namely, it will continue the proliferation of copies of

20   Licensed Materials divorced from their licenses and identifying information until infringement is

21   so prevalent no amount of enforcement by Plaintiffs and the Class could stop its spread.

22   **COUNT V**
23   **FALSE DESIGNATION OF ORIGIN—REVERSE PASSING OFF**
                              15 U.S.C. § 1125
24   **(GitHub and OpenAI)**

25   196.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding

26   and succeeding paragraph as though fully set forth herein.

27   197.   GitHub and OpenAI have used or made, and will continue to use or make, in

28   commerce throughout the United States, including in California, one or more words, terms,

---

names, symbols, or devices, or any combination thereof, or any false and/or misleading

designation of origin, false and/or misleading description of fact, or false and/or misleading

representation of fact that is likely to cause consumer confusion, or to cause mistake, or to deceive

as to the affiliation, connection, or association of Plaintiffs' and the Class's Licensed Materials

and Copilot, or as to the origin, sponsorship, or approval of Plaintiffs' and the Class's Licensed

Materials and Copilot.

198.   As a result, GitHub and OpenAI have intentionally violated 15 U.S.C. §

1125(a)(1)(A).

199.   As an actual and proximate result of GitHub's and OpenAI's acts, Plaintiffs and

the Class have suffered and continue to suffer harm.

COUNT VI
UNJUST ENRICHMENT
Cal. Bus. & Prof. Code §§ 17200, *et seq.* and Common Law
(GitHub and OpenAI)

200.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding

and succeeding paragraph as though fully set forth herein.

201.   Plaintiffs and the Class have invested substantial time and energy in creating the

Licensed Materials.

202.   GitHub and OpenAI have unjustly utilized access to Licensed Materials hosted on

GitHub. This code is used to create Derivative Works that are licensed to third parties in

exchange for, *inter alia*, compliance with applicable License terms.

203.   GitHub and OpenAI derive profit or other benefits from removal of attribution,

copyright notices, and license terms from Licensed Materials and reselling it as Output through

Copilot.

204.   It would be unjust for GitHub and OpenAI to retain those benefits.

205.   Plaintiffs and the Class have suffered monetary damages as a result of GitHub's

and OpenAI's conduct.

206.   The conduct of GitHub and OpenAI is causing and, unless enjoined and restrained by this Court, will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully be compensated or measured in money.

COUNT VII
UNFAIR COMPETITION
15 U.S.C. § 1125; *Cal. Bus. & Prof. Code* **§§ 17200,** *et seq.***; and Common Law**
(GitHub and OpenAI)

207.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

208.   GitHub and OpenAI have engaged in unlawful business practices, including:

a.   Violations of Plaintiffs' and the Class's rights under the DMCA;

b.   Tortious interference in Plaintiffs' and the Class's contractual relations with users of their code;

c.   Passing off Copilot's Output as originating from Copilot, GitHub, and/or OpenAI;

d.   Privacy violations, namely violation of the CCPA and negligent handling of personal information.

The details of the unlawful business practices are set forth herein.

209.   Plaintiffs and the Class have suffered monetary damages as a result of GitHub's and OpenAI's conduct.

210.   The conduct of GitHub and OpenAI is causing and, unless enjoined and restrained by this Court, will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully be compensated or measured in money.

COUNT VIII
BREACH OF CONTRACT
VIOLATION OF GITHUB PRIVACY POLICY and TERMS OF SERVICE
Cal. Bus. & Prof. Code **§ 22575–22579;** *Cal. Civ. Code* **§ 1798.150; and Common Law**
(Against GitHub)

211.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

212.   Plaintiffs and the Class are GitHub users who have accepted GitHub's Terms of Service. As a result, Plaintiffs and the Class have formed a contract, the terms of which are set forth in GitHub's Terms of Service—including the additional GitHub Copilot Terms from GitHub Terms for Additional Products and Features.

213.   Plaintiffs and the Class are GitHub users who have accepted GitHub's Privacy Statement. As a result, Plaintiffs and the Class have formed a contract.

214.   GitHub's Privacy Statement, Terms of Service, and GitHub Copilot Terms share definitions and refer to each other. As such, they are collectively referred to herein as "GitHub's Policies" unless a distinction is necessary and are attached as Exhibit 1.

215.   Plaintiffs and the Class have performed each of the conditions, covenants, and obligations imposed on them by the terms of GitHub's Policies.

216.   GitHub has substantially and materially breached GitHub's Policies in the following ways:

    a. Sharing Plaintiffs' and the Class's personal data with unauthorized third parties in violation of the GitHub Privacy Statement;

    b. Selling and distributing Plaintiffs' and the Class's personal data in contravention of the GitHub Policies;

    c. Use of Plaintiffs' and the Class's personal data after the GitHub Privacy Statement explicitly claims it will be deleted;

    d. Use and distribution of Plaintiffs' and the Class's personal data outside the limitations set forth in the GitHub Privacy Statement.

217.   Plaintiffs and the Class have suffered monetary damages as a result of GitHub's conduct.

218.   GitHub's conduct is causing and, unless enjoined and restrained by this Court, will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully be compensated or measured in money.

1 219. As a direct and proximate result of these material breaches by GitHub, Plaintiffs

2 and the Class are entitled to an injunction requiring GitHub to comply with all the terms of the

3 GitHub Policies.

4 220. Plaintiffs and the Class are further entitled to recover from GitHub the damages

5 Plaintiffs and the Class sustained—including consequential damages—for Plaintiffs' and the

6 Class's costs in enforcing GitHub's Policies. Plaintiffs and the Class are also entitled to recover as

7 restitution from GitHub for any unjust enrichment, including gains, profits, and advantages that it

8 has obtained as a result of its breaches of the GitHub Policies.

COUNT IX
VIOLATION OF THE CALIFORNIA CONSUMER PRIVACY ACT
Cal. Civ. Code **§ 1798.150**
(GitHub and OpenAI)

12 221. Plaintiffs and the Class hereby repeat and incorporate by reference each preceding

13 and succeeding paragraph as though fully set forth herein.

14 222. At all relevant times, GitHub and OpenAI were "businesses" under the terms of

15 Cal. Civ. Code § 1798.140(g) of the California Consumer Privacy Act ("CCPA") as sole

16 proprietorships, partnerships, limited liability companies, corporations, associations, or other

17 legal entities operating in the State of California that collect consumers' personal information.

18 GitHub and OpenAI either have annual operating revenue above $25 million, collect the personal

19 information of 50,000 or more California residents annually, or derive at least fifty percent of

20 their annual revenue from the sale of personal information of California residents.

21 223. At all relevant times, Plaintiffs and the Class were "consumers" under the terms

22 of the CCPA as natural persons as defined in Section 17014 of Title 18 of the California Code of

23 Regulations.

24 224. "Personal Information" is defined in Section 1798.140(o)(1) of the CCPA, which

25 protects consumers' personal information from collection, use, or sale by businesses without

26 consumers' notice and consent.

27 225. GitHub and OpenAI violated the CCPA by using customers' PII without providing

28 the required notice under the CCPA. GitHub and OpenAI did not notify Plaintiffs nor the Class

they were using, distributing, or selling their PII to unauthorized third parties, namely Copilot users.

226.   GitHub and OpenAI also violated the CCPA by failing to provide notice to its customers of their right to opt-out of the disclosure of their PII to unauthorized third parties, namely Copilot users.

227.   GitHub and OpenAI also violated the CCPA by incorporating Plaintiffs' and the Class's personal information into Copilot with no way to alter or delete. And also with no way to share that personal data with Plaintiffs or the Class upon request.

228.   GitHub and OpenAI also violated the CCPA by failing to provide a clear and conspicuous link entitled "Do Not Sell My Personal Information" to a webpage that enables a consumer—or a person authorized by a consumer—to opt out of the sale of Plaintiffs' and the Class's personal data through Copilot.

229.   By the acts described above, GitHub and OpenAI violated the CCPA by negligently, carelessly, and recklessly collecting, maintaining, and controlling their customers' sensitive personal information and by engineering, designing, maintaining, and controlling systems that exposed their customers' sensitive personal information of which GitHub and OpenAI had control and possession to the risk of exposure to unauthorized persons, thereby violating their duty to implement and maintain reasonable security procedures and practices appropriate to the nature of the information to protect the personal information. GitHub and OpenAI allowed unauthorized users to view, use, manipulate, exfiltrate, and steal the nonencrypted and nonredacted personal information of Plaintiffs and other customers, including their personal and financial information.

COUNT X
NEGLIGENCE—NEGLIGENT HANDLING OF PERSONAL DATA
Common Law
(GitHub and OpenAI)

230.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

231.   GitHub and OpenAI owed a duty of reasonable care toward Plaintiffs and the

Class based upon GitHub's and OpenAI's relationship to them. This duty is based upon

GitHub's and OpenAI's contractual obligations, custom and practice, right to control information

in its possession, exercise of control over the information in its possession, authority to control

the information in its possession, and the commission of affirmative acts that resulted in said

harms and losses. Additionally, this duty is based on the requirements of California Civil Code

section 1714 requiring all "persons," including GitHub and OpenAI, to act in a reasonable

manner toward others. This duty is also based on the specific statutory duties imposed on

GitHub and OpenAI under California Civil Code sections 1798.100, *et seq.*, as businesses

operating in the State of California that either have annual operating revenue above $25 million,

collect the personal information of 50,000 or more California residents annually, or derive at least

50 percent of their annual revenue from the sale of personal information of California residents.

232.   GitHub and OpenAI breached their duties by negligently, carelessly, and recklessly

collecting, maintaining, and controlling their customers' sensitive personal information and

engineering, designing, maintaining, and controlling systems—including Copilot—that exposed

and continue to expose their customers' sensitive personal information of which GitHub and

OpenAI had control and possession to the risk of exposure to unauthorized persons.

233.   GitHub and OpenAI also committed per se breaches of said duty by negligently

violating the dictates of California Civil Code sections 1798.82, *et seq.*, and 1798.100, *et seq.*, and

the provisions of the California Constitution enshrining the right to privacy, by failing to inform

Plaintiffs and the Class of the access to their sensitive personal information by unauthorized

persons expeditiously and without delay and failing to adequately safeguard this information from

unauthorized access even after GitHub and OpenAI became aware of multiple instances of

release of this information by Copilot. The provisions of the California Civil Code and the

California Constitution that GitHub and OpenAI violated were enacted to protect the class of

Plaintiffs here involved from the type of injury here incurred, namely their right to privacy and

the protection of their personal data. Plaintiffs and the Class were within the class of persons and

consumers who were intended to be protected by California Civil Code sections 1798.82, *et seq.*, and 1798.100, *et seq.*

234.   As a direct consequence of the actions described herein, and the breaches of duties indicated thereby, unauthorized users gained access to, exfiltrated, stole, and gained disclosure of the sensitive personal information of Plaintiffs and the Class, causing them harms and losses including but not limited to economic loss, the loss of control over the use of their identity, harm to their constitutional right to privacy, lost time dedicated to cure harm to their privacy, the need for future expenses and time dedicated to the recovery and protection of further loss, and privacy injuries associated with having their sensitive personal and financial information disclosed.

COUNT XI
CIVIL CONSPIRACY
Common Law
(Against All Defendants)

235.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

236.   On information and belief, Microsoft, GitHub, OpenAI, and the Individual Defendants have worked together to create Copilot. In creating Copilot, Defendants willfully avoided determining whether and how Copilot's training and Output may violate the rights of Plaintiffs and the Class and other stakeholders. This is because Defendants understood that through Copilot they would be engaging in a variety of unlawful conduct. Defendants conduct resulted in violations of Plaintiffs' and the Class's rights as set forth herein.

237.   On information and belief, OpenAI derives a financial or other valuable benefit from the sale of Copilot. In exchange, OpenAI provided Microsoft an exclusive license to use its GPT-3 language model.

238.   On information and belief, Microsoft derives a financial benefit from sales of Copilot through payments or other form of compensation in exchange for GitHub's and OpenAI's use of Azure to run Copilot.

239.   GitHub, Microsoft, and OpenAI agreed to a common plan or design to create, sell, and run Copilot to commit and conceal the following tortious acts:

    a.   Violations of Plaintiffs' and the Class's rights under the DMCA Section 1202;

    b.   Tortious interference in Plaintiffs' and the Class's contractual relations with users of their code;

    c.   Passing off Copilot's Output as originating either from Copilot itself or from Defendants;

    d.   Unfair competition with Plaintiffs and the Class by releasing and operating Copilot; and

    e.   Privacy violations, namely violation of the CCPA and negligent handling of personal information.

COUNT XII
DECLARATORY RELIEF
28 U.S.C. § 2201(a) and Cal. Code Civ. Proc. § 1060
(Against All Defendants)

240.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

241.   An actual controversy exists between the Class (including Plaintiffs) and Defendants due to Defendants' operation of Copilot, which violates Plaintiffs' and the Class's rights, including but not limited to their rights under DMCA Section 1202, the Lanham Act, the CCPA, and Cal. Civ. Code § 17200 as alleged herein.

242.   Plaintiffs are entitled to a judgment declaring that Defendants' actions are unlawful and, specifically, that Defendants violated DMCA Section 1202, the Lanham Act, the CCPA, and Cal. Civ. Code § 17200.

IX.   DEMAND FOR JUDGMENT

WHEREFORE, Plaintiffs requests that the Court enter judgment on their behalf and on behalf of the Class defined herein, by adjudging and decreeing that:

243.   This action may proceed as a class action, with Plaintiffs serving as Class Representatives, and with Plaintiffs' counsel as Class Counsel;

a) Judgment in favor of Plaintiffs and the Class and against Defendants;

b) Permanent injunctive relief, including but not limited to making changes to its Copilot product to ensure that all applicable information set forth in 17 U.S.C. § 1203(b)(1) is included in along with any Output including associated code;

c) An order of costs and allowable attorney's fees pursuant to 17 U.S.C. § 1203(b)(4)–(5);

d) An award of statutory damages pursuant to 17 U.S.C. § 1203(b)(3) and 17 U.S.C. § 1203(c)(3),[41] or, in the alternative, an award of actual damages and any additional profits pursuant to 17 U.S.C. § 1203(c)(2) (including tripling damages pursuant to 17 U.S.C. § 1203(c)(4) if applicable);

e) An award of damages for harms resulting from Defendants' breach of Licenses;

f) An award of damages, including punitive damages, for harms resulting from Defendants' tortious interference in Plaintiffs' and the Class's prospective contractual relations;

g) An award of damages for harms resulting from Defendants' false designation of the origin of Copilot's Output;

h) An award of damages in the amount Defendants have been unjustly enriched through their conduct as alleged herein as well as punitive damages in connection with this conduct;

i) An award of damages, including punitive damages, for harms resulting from Defendants acts of unfair competition;

j) Statutory damages and any other relief this Court deems proper for Defendants

---

[41] Plaintiffs estimate that statutory damages for Defendants' direct violations of DMCA Section 1202 alone will exceed $9,000,000,000. That figure represents minimum statutory damages ($2,500) incurred three times for each of the 1.2 million Copilot users Microsoft reported in June 2022. Each time Copilot provides an unlawful Output it violates Section 1202 three times (distributing the Licensed Materials without: (1) attribution, (2) copyright notice, and (3) License Terms). So, if each user receives just one Output that violates Section 1202 throughout their time using Copilot (up to fifteen months for the earliest adopters), then GitHub and OpenAI have violated the DMCA 3,600,000 times. At minimum statutory damages of $2500 per violation, that translates to $9,000,000,000.

1   violation of the CCPA;

2   k) An award of damages for harms resulting from GitHub's breach of the GitHub

3   Policies; and

4   l) An award of damages, including punitive damages, for harms resulting from

5   Defendants' negligent handling of Plaintiffs' and the Class's personal data.

6   244.   Injunctive relief sufficient to alleviate and stop Defendants' unlawful conduct

7 alleged herein.

8   245.   Plaintiffs and the Class are entitled to prejudgment and post-judgment interest on

9 the damages awarded them, and that such interest be awarded at the highest legal rate from and

10 after the date this class action complaint is first served on Defendants;

11   246.   Defendants are to be jointly and severally responsible financially for the costs and

12 expenses of a Court approved notice program through post and media designed to give immediate

13 notification to the Class.

14   247.   Plaintiffs and the Class receive such other or further relief as may be just and

15 proper.

16   **X.   JURY TRIAL DEMANDED**

17   Pursuant to Federal Rule of Civil Procedure 38(b), Plaintiffs demand a trial by jury of all

18 the claims asserted in this Complaint so triable.

19   / / /

20   / / /

21   / / /

22

23

24

25

26

27

28

1   Dated: November 3, 2022   By: */s/ Joseph R. Saveri*
                                                            Joseph R. Saveri
2

3

4   Joseph R. Saveri (State Bar No. 130064)
                                                 Cadio Zirpoli (State Bar No. 179108)
5   Travis Manfredi (State Bar No. 281779)
                                                 JOSEPH SAVERI LAW FIRM, LLP
6   601 California Street, Suite 1000
                                                 San Francisco, California 94108
7   Telephone: (415) 500-6800
                                                 Facsimile: (415) 395-9940
8   Email: jsaveri@saverilawfirm.com
                                                               czirpoli@saverilawfirm.com
9   tmanfredi@saverilawfirm.com
10

11

12   Matthew Butterick (State Bar No. 250953)
                                                 1920 Hillhurst Avenue, #406
13   Los Angeles, CA 90027
                                                 Telephone: (323) 968-2632
14   Facsimile: (415) 395-9940
                                                 Email: mb@butterricklaw.com
15
                                                 Counsel for Plaintiffs and the Proposed Class
16

17

18

19

20

21

22

23

24

25

26

27

28

EXHIBIT D

Joseph R. Saveri (State Bar No. 130064)
Cadio Zirpoli (State Bar No. 179108)
Christopher K.L. Young (State Bar No. 318371)
Louis A. Kessler (State Bar No. 243703)
Elissa A. Buchanan (State Bar No. 249996)
Travis Manfredi (State Bar No. 281779)
William W. Castillo Guardado (State Bar No. 294159)
Holden J. Benon (State Bar No. 325847)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1000
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
                czirpoli@saverilawfirm.com
                cyoung@saverilawfirm.com
                lkessler@saverilawfirm.com
                eabuchanan@saverilawfirm.com
                tmanfredi@saverilawfirm.com
                wcastillo@saverilawfirm.com
                hbenon@saverilawfirm.com

*Counsel for Individual and Representative
Plaintiffs and the Proposed Class*

[Additional Counsel Listed on Signature Page]

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
OAKLAND DIVISION

J. DOE 1, J. DOE 2, J. DOE 3, J. DOE 4, and J. DOE 5, individually and on behalf of all others similarly situated, *Individual and Representative Plaintiffs,* v. GITHUB, INC., a Delaware corporation; MICROSOFT CORPORATION, a Washington corporation; OPENAI, INC., a Delaware nonprofit corporation; OPENAI, L.P., a Delaware limited partnership; OPENAI OPCO, L.L.C., a Delaware limited liability company; OPENAI GP, L.L.C., a Delaware limited liability company; OPENAI STARTUP FUND GP I, L.L.C., a Delaware limited liability company; OPENAI STARTUP FUND I, L.P., a Delaware limited partnership; OPENAI STARTUP FUND MANAGEMENT, LLC, a Delaware limited liability company; OPENAI, L.L.C. , a Delaware limited liability company; OPENAI GLOBAL, LLC , a Delaware limited liability company; OAI CORPORATION, a Delaware corporation; OPENAI HOLDINGS, LLC, a Delaware limited liability company; OPENAI HOLDCO, LLC, a Delaware limited liability company; OPENAI INVESTMENT	Case No.: 4:22-cv-06823-JST 4:22-cv-07074-JST **SECOND AMENDED COMPLAINT** **CLASS ACTION** **DEMAND FOR JURY TRIAL** ***FILED UNDER SEAL*** **[CONFIDENTIAL]**

Case No. 4:22-cv-06823-JST

1   LLC, a Delaware limited liability company; OPENAI
    STARTUP FUND SPV I, L.P. , a Delaware limited
2   partnership; and OPENAI STARTUP FUND SPV GP
    I, L.L.C. , a Delaware limited liability company;

3   *Defendants.*

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

TABLE OF CONTENTS

SECOND AMENDED COMPLAINT

SECOND AMENDED COMPLAINT

Plaintiffs J. Doe 1, J. Doe 2, J. Doe 3, J. Doe 4 and J. Doe 5 ("Plaintiffs"), on behalf of themselves and all others similarly situated, bring this Class Action Complaint (the "Complaint") against Defendants GitHub, Inc.; Microsoft Corporation; OpenAI, Inc.; OpenAI, L.P.; OpenAI OpCo, L.L.C.; OpenAI GP, L.L.C.; OpenAI Startup Fund GP I, L.L.C.; OpenAI Startup Fund I, L.P.; OpenAI Startup Fund Management, LLC; OpenAI, L.L.C.; OpenAI Global, LLC ("OpenAI Global"); OAI Corporation ("OAI"); OpenAI Holdings, LLC ("OpenAI Holdings"); OpenAI Holdco, LLC; OpenAI Investment L.L.C.; OpenAI Startup Fund SPV I, L.P.; and OpenAI Startup Fund SPV GP I, LLC[1] for violation of the Digital Millennium Copyright Act, 17 U.S.C. §§ 1201–1205 (the "DMCA"); breach of contract regarding the Suggested Licenses, and breach of contract regarding GitHub's policies including its terms of service.

I.   OVERVIEW: A BRAVE NEW WORLD
OF SOFTWARE PIRACY

1.   Plaintiffs and Class members are owners of copyright interests in materials made available publicly on GitHub that are subject to various licenses containing conditions for use of those works (the "Licensed Materials"). All the licenses at issue here (the "Licenses") contain certain common terms (the "License Terms").

2.   "Artificial Intelligence" is referred to herein as "AI." AI is defined for the purposes of this Complaint as a computer program that algorithmically simulates human reasoning or inference, often using statistical methods. Machine Learning ("ML") is a subset of AI in which the behavior of the program is derived from analyzing a corpus of material called training data.

3.   GitHub is a company founded in 2008 by a team of open-source enthusiasts. At the time, GitHub's stated goal was to support open-source development, especially by hosting open-source source

---

[1] GitHub, Inc. is referred to as "GitHub." Microsoft Corporation is referred to as "Microsoft." OpenAI, Inc.; OpenAI, L.P.; OpenAI OpCo, L.L.C.; OpenAI GP, L.L.C.; OpenAI Startup Fund GP I, L.L.C.; OpenAI Startup Fund I, L.P.; OpenAI Startup Fund Management, LLC; OpenAI, L.L.C.; OpenAI Global; OAI Corporation ("OAI"); OpenAI Holdings; OpenAI Holdco, LLC; OpenAI Investment LLC; OpenAI Startup Fund SPV I, L.P.; and OpenAI Startup Fund SPV GP I, L.L.C. are referred to collectively herein as "OpenAI." Collectively, GitHub, Inc., Microsoft Corporation, OpenAI, Inc.; OpenAI, L.P.; OpenAI GP, L.L.C.; OpenAI Startup Fund GP I, L.L.C.; OpenAI Startup Fund I, L.P.; OpenAI Startup Fund Management, LLC; OpenAI, L.L.C.; OpenAI Global; OAI; OpenAI Holdings; OpenAI Holdco, LLC; OpenAI Investment LLC; OpenAI Startup Fund SPV I, L.P.; and OpenAI Startup Fund SPV GP I, L.L.C. are referred to herein as "Defendants."

Case No. 4:22-cv-06823-JST   1
SECOND AMENDED COMPLAINT

1 code on the website github.com. Over the next 10 years, GitHub, based on these representations

2 succeeded wildly, attracting nearly 25 million developers.

3   4.   Developers published Licensed Materials on GitHub pursuant to written Licenses. In

4 particular, the most popular ones share a common term: use of the Licensed Materials requires some form

5 of *attribution*, usually by, among other things, including a copy of the license along with the name and

6 copyright notice of the original author.

7   5.   On October 26, 2018, Microsoft acquired GitHub for $7.5 billion. Though some members

8 of the open-source community were skeptical of this union, Microsoft repeated one mantra throughout:

9 "Microsoft Loves Open Source." For the first few years, Microsoft's representations seemed credible.

10   6.   Microsoft invested $1 billion in OpenAI LP in July 2019 at a $20 billion valuation. In 2020,

11 Microsoft became exclusive licensee of OpenAI's GPT-3 language model—despite OpenAI's continued

12 claims its products are meant to benefit "humanity" at large. In 2021, Microsoft began offering GPT-3

13 through its Azure cloud-computing platform. On October 20, 2022, it was reported that OpenAI "is in

14 advanced talks to raise more funding from Microsoft" at that same $20 billion valuation. Copilot runs on

15 Microsoft's Azure platform. Microsoft has used Copilot to promote Azure's processing power,

16 particularly regarding AI.

17   7.   On information and belief, Microsoft obtained a partial ownership interest in OpenAI in

18 exchange for its $1 billion investment. As OpenAI's largest investor and largest service provider—

19 specifically in connection with Microsoft's Azure product—Microsoft exerts considerable control over

20 OpenAI.

21   8.   In June 2021, GitHub and OpenAI launched Copilot, an AI-based product that promises to

22 assist software coders by providing or filling in blocks of code using AI. GitHub charges Copilot users $10

23 per month or $100 per year for this service. Copilot ignores, violates, and removes the Licenses offered by

24 thousands—possibly millions—of software developers, thereby accomplishing software piracy on an

25 unprecedented scale. Copilot outputs text derived from Plaintiffs' and the Class's Licensed Materials

26 without adhering to the applicable License Terms and applicable laws. Copilot's output is referred herein

27 as "Output."

28

9.   On August 10, 2021, OpenAI debuted its Codex product, which converts natural language into code and is integrated into Copilot. Copilot and Codex can be called either AIs or MLs. Codex and Copilot will be referred to as Ais herein unless a distinction is required.

10.   Though Defendants have been cagey about what data was used to train the AI,[2] they have conceded that the training data includes data in vast numbers of publicly accessible repositories on GitHub,[3] which include and are limited by Licenses.

11.   Among other things, Defendants stripped Plaintiffs' and the Class's attribution, copyright notice, and license terms from their code in violation of the Licenses and Plaintiffs' and the Class's rights. Defendants used Copilot to distribute the now-anonymized code to Copilot users as if it were created by Copilot.

12.   Copilot is run entirely on Microsoft's Azure cloud-computing platform.

13.   Copilot often simply reproduces code that can be traced back to open-source repositories or open-source licensees. Contrary to and in violation of the Licenses, code reproduced by Copilot *never* includes attributions to the underlying authors.

14.   GitHub and OpenAI have offered shifting accounts of the source and amount of the code or other data used to train and operate Copilot. They have also offered shifting justifications for why a commercial AI product like Copilot should be exempt from these license requirements, often citing "fair use."

15.   It is not fair, permitted, or justified. On the contrary, Copilot's goal is to replace a huge swath of open source by taking it and keeping it inside a GitHub-controlled paywall. It violates the licenses that open-source programmers chose and monetizes their code despite GitHub's pledge never to do so.

---

[2] "Training" an AI, as described in greater detail below, means feeding it large amounts of data that it interprets using given criteria. Feedback is then given to it to fine-tune its Output until it can provide Output with minimal errors.

[3] Repositories are containers for individual coding projects. They are where GitHub users upload their code and where other users can find it. Most GitHub users have multiple repositories.

[CONFIDENTIAL]

II. JURISDICTION AND VENUE

16.   Plaintiffs bring this action on their own behalf as well as representatives of a Class of similarly situated individuals and entities. They seek to recover injunctive relief and damages as a result and consequence of Defendants' unlawful conduct.

17.   Jurisdiction and venue are proper in this judicial district under 28 U.S.C. § 1331 pursuant to Defendants' violation of Section 1202(b) of the Digital Millennium Copyright Act, 17 U.S.C. §§ 1201–1205; and because a substantial part of the events giving rise to Plaintiffs' claims occurred in this District, a substantial portion of the affected interstate trade and commerce was carried out in this District, and three or more of the Defendants reside in this District and/or are licensed to do business in this District. Each Defendant has transacted business, maintained substantial contacts, and/or committed overt acts in furtherance of the illegal scheme and conspiracy throughout the United States, including in this District. Defendants' conduct has had the intended and foreseeable effect of causing injury to persons residing in, located in, or doing business throughout the United States, including in this District.

III. INTRADISTRICT ASSIGNMENT

18.   Pursuant to Civil Local Rule 3.2 (c) and (e), assignment of this case to the San Francisco Division of the United States District Court for the Northern District of California is proper because a substantial amount of the development of the Copilot product as well as of the interstate trade and commerce involved and affected by Defendants' conduct giving rise to the claims herein occurred in this Division. Furthermore, Defendants GitHub and all the OpenAI entities are headquartered within this Division.

IV. PARTIES

A. Plaintiffs

19.   Plaintiff J. Doe 1, ███████, is a resident of the State of New Hampshire. Plaintiff Doe 1 published Licensed Materials they owned a copyright interest in to at least one GitHub repository under one of the Suggested Licenses. Specifically, Doe 1 has published Licensed Materials they claim a copyright interest in under the following Suggested Licenses: MIT License and GNU General Public

[CONFIDENTIAL]

1 License version 3.0. Plaintiff was, and continues to be, injured during the Class Period as a result of

2 Defendants' unlawful conduct alleged herein.

3   20.   Plaintiff J. Doe 2, ███████, is a resident of the State of Illinois. Plaintiff Doe 2 published

4 Licensed Materials they owned a copyright interest in to at least one GitHub repository under one of the

5 Suggested Licenses. Specifically, Doe 2 has published Licensed Materials they claim a copyright interest

6 in under the following Suggested Licenses: MIT License; GNU General Public License version 3.0; GNU

7 Affero General Public License version 3.0; The 3-Clause BSD License; and Apache License 2.0. Plaintiff

8 was, and continues to be, injured during the Class Period as a result of Defendants' unlawful conduct

9 alleged herein.

10   21.   Plaintiff J. Doe 3, ████████, is a resident of the State of Idaho. Plaintiff Doe 3

11 published Licensed Materials they owned a copyright interest in to at least one GitHub repository under

12 one of the Suggested Licenses. Specifically, Doe 3 has published Licensed Materials they claim a

13 copyright interest in under the following Suggested Licenses: MIT License; GNU General Public License

14 version 3.0; and GNU Affero General Public License version 3.0. Plaintiff was, and continues to be,

15 injured during the Class Period as a result of Defendants' unlawful conduct alleged herein.

16   22.   Plaintiff J. Doe 4, ████████, is a resident of the State of South Carolina. Plaintiff Doe 4

17 published Licensed Materials they owned a copyright interest in to at least one GitHub repository under

18 one of the Suggested Licenses. Specifically, Doe 4 has published Licensed Materials they claim a

19 copyright interest in under the following Suggested Licenses: GNU General Public License v2.0 and

20 GNU General Public License v3.0. Plaintiff was, and continues to be, injured during the Class Period as a

21 result of Defendants' unlawful conduct alleged herein.

22   23.   Plaintiff J. Doe 5, ████████, is a resident of the Commonwealth of Massachusetts.

23 Plaintiff Doe 5 published Licensed Materials they owned a copyright interest in to at least one GitHub

24 repository under one of the Suggested Licenses. Specifically, Doe 5 has published Licensed Materials they

25 claim a copyright interest in under the following Suggested Licenses: MIT License; Apache License 2.0;

26 and GNU General Public License v3.0.

27

28

B. Defendants

24.   Defendant GitHub, Inc. is a Delaware corporation with its principal place of business located at 88 Colin P Kelly Jr Street, San Francisco, CA 94107. GitHub sells, markets, and distributes Copilot throughout the internet and other sales channels throughout the United States, including in this District. GitHub released Copilot on a limited "technical preview" basis on June 29, 2021. On June 21, 2022, Copilot was released to the public as a subscription-based service for individual developers. GitHub is a party to the unlawful conduct alleged herein.

25.   Defendant Microsoft Corporation is a Washington corporation with its principal place of business located at One Microsoft Way, Redmond, Washington 98052. Microsoft announced its acquisition of Defendant GitHub, Inc. on June 4, 2018. On October 26, 2018, Microsoft finalized its acquisition of GitHub. Microsoft owns and operates GitHub. Through its corporate ownership, control of the GitHub Board of Directors, active management, and other means, Microsoft sells, markets, and distributes Copilot. Microsoft is a party to the unlawful conduct alleged herein.

26.   Defendant OpenAI, Inc. is a Delaware nonprofit corporation with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI, Inc. is a party to the unlawful conduct alleged herein. It—along with OpenAI, L.P.—programed, trained, and maintains Codex, which infringes all the same rights at Copilot and is also an integral piece of Copilot. Copilot requires Codex to function. OpenAI, Inc. is a party to the unlawful conduct alleged herein. OpenAI, Inc. founded, owns, and exercises control over all the other OpenAI entities, including those set forth in Paragraphs 27–40.

27.   Defendant OpenAI, L.P. is a Delaware limited partnership with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI, L.P. is a party to the unlawful conduct alleged herein. Its primary activity is research and technology. OpenAI, L.P. is a wholly owned subsidiary of OpenAI, Inc. that is operated for profit. OpenAI, L.P. is the OpenAI entity that co-created Copilot and offers it jointly with GitHub. OpenAI's revenue, including revenue from Copilot, is received by OpenAI, L.P. OpenAI, Inc. controls OpenAI, L.P. directly and through the other OpenAI entities.

28.   Defendant OpenAI OpCo, L.L.C. is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI OpCo, L.L.C. is a party to the unlawful conduct alleged herein. Its primary activity is research and technology. OpenAI OpCo,

L.L.C. is a wholly owned subsidiary of OpenAI, Inc. that is operated for profit. OpenAI OpCo, L.L.C. is the OpenAI entity that co-created Copilot and offers it jointly with GitHub. OpenAI's revenue, including revenue from Copilot, is received by OpenAI OpCo, L.L.C. OpenAI, Inc. controls OpenAI OpCo, L.L.C. directly and through the other OpenAI entities.

29.   Defendant OpenAI GP, L.L.C. ("OpenAI GP") is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI GP is the general partner of OpenAI, L.P. OpenAI GP manages and operates the day-to-day business and affairs of OpenAI, L.P. OpenAI GP is liable for the debts, liabilities and obligations of OpenAI, L.P., including litigation and judgments. OpenAI GP is a party to the unlawful conduct alleged herein. Its primary activity is research and technology. OpenAI GP is the general partner of OpenAI, L.P. OpenAI GP was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period. OpenAI, Inc. directly controls OpenAI GP. OpenAI GP directly controls OpenAI Holdings and OpenAI Global.

30.   Defendant OpenAI Startup Fund I, L.P. ("OpenAI Startup Fund I") is a Delaware limited partnership with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI Startup Fund I was instrumental in the foundation of OpenAI, L.P., including the creation of its business strategy and providing initial funding. Through participation in OpenAI Startup Fund I, certain entities and individuals obtained an ownership interest in OpenAI, L.P. Plaintiffs are informed and believed, and on that basis allege that OpenAI Startup Fund I participated in the organization and operation of OpenAI, L.P. OpenAI Startup Fund I is a party to the unlawful conduct alleged herein. OpenAI Startup Fund I was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period.

31.   Defendant OpenAI Startup Fund GP I, L.L.C. ("OpenAI Startup Fund GP I") is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI Startup Fund GP I is the general partner of OpenAI Startup Fund I. OpenAI Startup Fund GP I manages and operates the day-to-day business and affairs of OpenAI Startup Fund I. OpenAI Startup Fund GP I is liable for the debts, liabilities and obligations of OpenAI Startup Fund I, including litigation and judgments. OpenAI Startup Fund GP I was aware of the unlawful conduct

alleged herein and exercised control over OpenAI, L.P. throughout the Class Period. OpenAI Startup Fund GP I is a party to the unlawful conduct alleged herein. Sam Altman, co-founder, CEO, and Board member of OpenAI, Inc. is the Manager of OpenAI Startup Fund GP I. OpenAI Startup Fund GP I is the General Partner of OpenAI Startup Fund I, L.P.

32.   Defendant OpenAI Startup Fund Management, LLC ("OpenAI Startup Fund Management") is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI Startup Fund Management is a party to the unlawful conduct alleged herein. OpenAI Startup Fund Management was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period.

33.   Defendant OpenAI, L.L.C. is a Delaware limited liability company with its principal place of business in San Francisco, California. OpenAI LLC owns some or all of the services and products provided by OpenAI. The sole member of OpenAI, L.L.C. is Defendant OpenAI OpCo, L.L.C.

34.   Defendant OpenAI Global, LLC is a Delaware limited liability company with its principal place of business in San Francisco, California. OpenAI Global's only members are Microsoft and Defendant OAI Corporation. Microsoft owns 49% of OpenAI Global, and exercises control over it as its largest minority shareholder. OpenAI describes OpenAI Global as a "capped profit company"),

35.   Defendant OAI Corporation ("OAI") is a Delaware corporation with its principal place of business in San Francisco, California. OAI's only member is Defendant OpenAI Holdings LLC.

36.   Defendant OpenAI Holdings, LLC is a Delaware limited liability company with its principal place of business in San Francisco, California. The members of OpenAI Holdings are Defendant OpenAI, Inc. and Aestas LLC, an OpenAI-related limited liability company that is not named as a defendant as of December 22, 2023. OpenAI Holdings is partially owned by OpenAI employees and outside investors.

37.   Defendant OpenAI Holdco, LLC is a Delaware limited liability company with its principal place of business in San Francisco, California.

38.   Defendant OpenAI Investment LLC is a Delaware limited liability company with its principal place of business in San Francisco, California.

39.   Defendant OpenAI Startup Fund SPV I, L.P. is a Delaware limited partnership with its principal place of business in San Francisco, California.

40.   Defendant OpenAI Startup Fund SPV GP I, L.L.C. is a Delaware limited liability company with its principal place of business in San Francisco, California. OpenAI Startup Fund SPV GP I, L.L.C. is the general partner and controls OpenAI Startup Fund SPV I, L.P.

V. AGENTS AND CO-CONSPIRATORS

41.   The unlawful acts alleged against the Defendants in this class action complaint were authorized, ordered, or performed by the Defendants' respective officers, agents, employees, representatives, or shareholders while actively engaged in the management, direction, or control of the Defendants' businesses or affairs.

42.   The Defendants' agents operated under the explicit and apparent authority of their principals.

43.   Each Defendant, and its subsidiaries, affiliates and agents operated as a single unified entity.

44.   Various persons and/or firms not named as Defendants herein may have participated as coconspirators in the violations alleged herein and may have performed acts and made statements in furtherance thereof.

45.   Each acted as the principal, agent, or joint venture of, or for other Defendants with respect to the acts, violations, and common course of conduct alleged herein.

VI. CLASS ALLEGATIONS

A. Class Definitions

46.   Plaintiffs bring this action for damages and injunctive relief on behalf of themselves and all others similarly situated as a class action pursuant to Rules 23(a), 23(b)(2), and 23(b)(3) of the Federal Rules of Civil Procedure, on behalf of the following Classes:

"Injunctive Relief Class" under Rule 23(b)(2):

> All persons or entities domiciled in the United States that, (1) owned an interest in at least one US copyright in any work; (2) offered that work under one of GitHub's Suggested Licenses[4]; and (3) stored Licensed Materials in any public GitHub repositories at any time between January 1, 2015 and the present (the "Class Period").

"Damages Class" under Rule 23(b)(3):

> All persons or entities domiciled in the United States that, (1) owned an interest in at least one US copyright in any work; (2) offered that work under one of GitHub's Suggested Licenses; and (3) stored Licensed Materials in any public GitHub repositories at any time during the Class Period.

These "Class Definitions" specifically exclude the following person or entities:

    a.   Any of the Defendants named herein;

    b.   Any of the Defendants' co-conspirators;

    c.   Any of Defendants' parent companies, subsidiaries, and affiliates;

    d.   Any of Defendants' officers, directors, management, employees, subsidiaries, affiliates, or agents;

    e.   All governmental entities; and

    f.   The judges and chambers staff in this case, as well as any members of their immediate families.

---

[4] When a GitHub user creates a new repository, they have the option of selecting one of thirteen licenses from a dropdown menu to apply to the contents of that repository. (They can also apply a different license later, or no license.) The Creative Commons Zero v1.0 Universal and the Unlicense donate the covered work to the public domain and/or otherwise waive all copyrights and related rights. Because they do not contain the necessary provisions nor do they even allow the owner to make copyright claims in most circumstances, they are not included in the Class Definition. We refer to the remaining eleven options as the "Suggested Licenses," which are: (1) Apache License 2.0 ("Apache 2.0"); (2) GNU General Public License version 3 ("GPL-3.0"); (3) MIT License ("MIT"); (4) The 2-Clause BSD License ("BSD 2"); (5) The 3-Clause BSD License ("BSD 3"); (6) Boost Software License ("BSL-1.0"); (7) Eclipse Public License 2.0 ("EPL-2.0"); (8) GNU Affero General Public License version 3 ("AGPL-3.0"); (9) GNU General Public License version 2 ("GPL-2.0"); (10) GNU Lesser General Public License version 2.1 ("LGPL-2.1"); and (11) Mozilla Public License 2.0 ("MPL-2.0"). These Suggested Licenses each contain at least three common requirements for use of the Licensed Materials in a derivative work or copy: attribution to the owner of the Licensed Materials ("Attribution"), inclusion of a copyright notice ("Copyright Notice"), and inclusion of the applicable Suggested License's text ("License Terms").

B. Numerosity

47.   Plaintiffs do not know the exact number of Class members, because such information is in the exclusive control of Defendants. Plaintiffs are informed and believe that there are at least thousands of Class members geographically dispersed throughout the United States such that joinder of all Class members in the prosecution of this action is impracticable.

C. Typicality

48.   Plaintiffs' claims are typical of the claims of their fellow Class members because Plaintiffs and Class members all own code published under a License. Plaintiffs and the Class published work subject to a License to GitHub later used by Copilot. Plaintiffs and absent Class members were damaged by this and other wrongful conduct of Defendants as alleged herein. Damages and the other relief sought herein is common to all members of the Class.

D. Commonality & Predominance

49.   Numerous questions of law or fact common to the entire Class arise from Defendants' conduct—including, but not limited to those identified below:

1.   DMCA Violations

- Whether Defendants' conduct violated the Class's rights under the DMCA when GitHub and OpenAI caused Codex and Copilot to ingest and distribute Licensed Materials without including any associated Attribution, Copyright Notice, or License Terms.

2.   Contract-Related Conduct

- Whether Defendants violated the Licenses governing use of the Licensed Materials by using them to train Copilot and for republishing those materials without appending the required Attribution, Copyright Notice, or License Terms.

- Whether Defendants interfered in prospective economic relations between the Class and the public regarding the Licensed Materials by concealing the License Terms.

- Whether Defendants intentionally or negligently interfered with a prospective economic advantage.

3.   Injunctive Relief

- Whether this Court should enjoin Defendants from engaging in the unlawful conduct alleged herein. And what the scope of that injunction would be.

4.   Defenses

- Whether any affirmative defense excuses Defendants' conduct.
- Whether any statutes of limitation limit Plaintiffs' and the Class's potential for recovery.
- Whether any applicable statutes of limitation should be tolled as a result of Defendants' fraudulent concealment of their unlawful conduct.

50.   These and other questions of law and fact are common to the Class and predominate over any questions affecting the Class members individually.

E. Adequacy

51.   Plaintiffs will fairly and adequately represent the interests of the Class because they have experienced the same harms as the Class and have no conflicts with any other members of the Class. Furthermore, Plaintiffs have retained sophisticated and competent counsel ("Class Counsel") who are experienced in prosecuting Federal and state class actions throughout the United States and other complex litigation and have extensive experience advising clients and litigating intellectual property, competition, contract, and privacy matters.

F. Other Class Considerations

52.   Defendants have acted on grounds generally applicable to the Class, thereby making final injunctive relief appropriate with respect to the Class as a whole.

53.   This class action is superior to alternatives, if any, for the fair and efficient adjudication of this controversy. Prosecuting the claims pleaded herein as a class action will eliminate the possibility of repetitive litigation. There will be no material difficulty in the management of this action as a class action.

54.   The prosecution of separate actions by individual Class members would create the risk of inconsistent or varying adjudications, establishing incompatible standards of conduct for Defendants.

VII.   FACTUAL ALLEGATIONS

A. Introduction

55.   This class action against Defendants concerns an OpenAI product called Codex and a GitHub product called Copilot.

56.   OpenAI began development of Codex sometime after OpenAI was founded in December 2015 and released Codex on a limited basis in August 2021.

57.   GitHub began development of Copilot sometime in 2019, released it on a limited basis in June 2021, and released it as a paid subscription service in June 2022.

58.   Codex and Copilot are assistive AI-based systems offered to software programmers. These AI systems are each trained on a large corpus of publicly accessible software code and other materials, including all the Licensed Materials. Defendants describe these products as making coding easier by accepting a code "prompt" from a programmer and emitting a possible completion of that code, which is referred to herein as Output, as set forth above. Defendants claim Codex and Copilot do not retain copies of the materials they are trained on. In practice, however, the Output is often a near-identical reproduction of code from the training data.

59.   Codex and Copilot are related. Copilot is a joint venture between GitHub and OpenAI. Codex is a standalone product released by OpenAI that also "powers GitHub Copilot, which [OpenAI] built and launched in partnership with GitHub."[5] "GitHub Copilot uses the OpenAI Codex to suggest code and entire functions in real-time, right from your editor."[6] The version of Codex used to power Copilot is modified from OpenAI's standalone version. Defendants have kept secret the details of Codex's modifications and its integration into or interaction with Copilot.

B. Codex Outputs Copyrighted Materials Without Following the Terms of the Applicable Licenses

60.   Below is an explanation of how Codex functions. When Codex is prompted with:

function isEven(n) {

---

[5] *See* https://openai.com/blog/openai-codex/.

[6] *See* https://github.com/features/copilot/.

it assumes this is the beginning of a function written in the JavaScript language that will test whether a number is even.

61.   Based on this assumption, Codex will then provide Output meant to complete the rest of the function. Based on the given prompt, it produced the following response:[7]

```
function isEven(n) {
  if (n == 0)
    return true;
  else if (n == 1)
    return false;
  else if (n < 0)
    return isEven(-n);
  else
    return isEven(n - 2);
}
console.log(isEven(50));
// → true
console.log(isEven(75));
// → false
console.log(isEven(-1));
// → ??
```

62.   The function itself occupies the first ten lines. Six additional lines follow the function, beginning with "console.log(isEven(50))". One possible explanation for Codex's inclusion of these lines is to test the "isEven" function. Though not part of the function itself, the lines will confirm the function works for certain values. In this case, the code implies that "isEven(50)" should return the value "true", and "isEven(75)" should return "false". Those answers are correct.

63.   The penultimate line indicates "isEven(-1)" should return "??". This is an error, as "isEven(-1)" should return "false".

64.   Codex cannot and does not understand the meaning of software code or any other Licensed Materials. But in training, what became Codex was exposed to an enormous amount of existing software code (its "Training Data") and—with input from its trainers and its own internal processes—

---

[7] Due to the nature of Codex, Copilot, and AI in general, Plaintiffs cannot be certain these examples would produce the same results if attempted following additional trainings of Codex and/or Copilot. However, these examples are representative of Codex and Copilot's Output at the time just prior to the filing of this Complaint.

inferred certain statistical patterns governing the structure of code and other Licensed Materials. The finished version of Codex, once trained, is known as a "Model."

65.   When given a prompt, such as the initial prompt discussed above—"function isEven(n) {"—Codex identifies the most statistically likely completion, based on the examples it reviewed in training. Every instance of Output from Codex is derived from material in its Training Data. Most of its Training Data consisted of Licensed Materials.

66.   Codex does not "write" code the way a human would, because it does not understand the meaning of code. Codex's lack of understanding of code is evidenced when it emits extra code that is not relevant under the circumstances. Here, Codex was only prompted to produce a function called "isEven". To produce its answer, Codex relied on Training Data that also appended the extra testing lines. Having encountered this function and the follow-up lines together frequently, Codex extrapolates they are all part of one function. A human with even a basic understanding of how JavaScript works would know the extra lines are not part of the function itself.

67.   Beyond the superfluous and inaccurate extra lines, this "isEven" function also contains two major defects. First, it assumes the variable "n" holds an integer. It could contain some other kind of value, like a decimal number or text string, which would cause an error. Second, even if "n" does hold an integer, the function will trigger a memory error called a "stack overflow" for sufficiently large integers. For these reasons, experienced programmers would not use Codex's Output.

68.   Codex does not identify the owner of the copyright to this Output, nor any other—it has not been trained to provide Attribution. Nor does it include a Copyright Notice nor any License Terms attached to the Output. This is by design—Codex was not coded or trained to track or reproduce such data. The Output in the example above is taken from *Eloquent JavaScript* by Marijn Haverbeke.[8]

69.   Here is the exercise from *Eloquent JavaScript*:

```
  // Your code here.
```

---

[8] https://eloquentjavascript.net/code/#3.2. *Eloquent JavaScript* is "Licensed under a Creative Commons [A]ttribution-[N]oncommercial license. All code in this book may also be considered licensed under an MIT license." *See* https://eloquentjavascript.net/. Thus, having also been posted on GitHub, the code Codex relied on meets the definition of Licensed Materials.

```
console.log(isEven(50));
// → true
console.log(isEven(75));
// → false
console.log(isEven(-1));
// → ??
```

70.   The exercise includes the "??" error. However, for Haverbeke's purposes, this is not an error but a placeholder value for the reader to fill in. Codex—as a mere probabilistic model—fails to recognize this nuance. The inclusion of the double question marks confirms unequivocally that Codex took this code directly from a copyrighted source without following any of the attendant License Terms.

71.   Haverbeke provides the following solution to the function discussed above:

```
function isEven(n) {
  if (n == 0) return true;
  else if (n == 1) return false;
  else if (n < 0) return isEven(-n);
  else return isEven(n - 2);
}

console.log(isEven(50));
// → true
console.log(isEven(75));
// → false
console.log(isEven(-1));
// → false
```

72.   Aside from different line breaks—which are not semantically meaningful in JavaScript—this code for the function "isEven" is the same as what Codex produced. The tests are also the same, though in this case Haverbeke provides the right answer for "isEven(-1)", which is "false". Codex has reproduced Haverbeke's Licensed Material almost verbatim, with the only difference being drawn from a different portion of those same Licensed Materials.

73.   There are many copies of Haverbeke's code stored in public repositories on GitHub, where programmers who are working through Haverbeke's book store their answers.

74.   The MIT license provides that "The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software."[9] Any person taking this code

---

[9] *See* Appendix A for full text of the MIT License.

1 directly from *Eloquent JavaScript* would have direct access to these License Terms and know to follow

2 them if incorporating the Licensed Materials into a derivative work and/or copying them. Codex does not

3 provide these License Terms.

4   75.   OpenAI Codex's Output would frequently, perhaps even constantly, contain Licensed

5 Materials, i.e., it would have conditions associated with it through its associated license. In its 2021

6 research paper about Codex called "Evaluating Large Language Models Trained on Code," OpenAI

7 stated Codex's Output is "often incorrect" and can contain security vulnerabilities and other

8 "misalignments" (meaning, departures from what the user requested).

9   76.   Most open-source licenses require attribution of the author, notice of their copyright, and a

10 copy of the license specifically to ensure that future coders can easily credit all previous authors and

11 ensure they adhere to all applicable licenses. All the Suggested Licenses include these requirements.

12   77.   Ultimately, Codex derives its value primarily from its ability to locate and output

13 potentially useful Licensed Materials. And from its obfuscation of any rights associated with those

14 materials.

15 **C. Copilot Outputs Copyrighted Materials Without Following the Terms of the Applicable**
   Licenses
16

17   78.   GitHub Copilot works in a similar way to OpenAI Codex. As mentioned above, a modified

18 version of Codex is used as the engine that powers Copilot.

19   79.   Copilot is installed by the end user as an extension to various code editors, including

20 Microsoft's Visual Studio and VS Code. As the user types into the editor, their code is uploaded in real

21 time to Microsoft's Azure cloud platform, where they become prompts for Copilot.

22   80.   When we give Copilot the same prompt discussed above in Paragraph 61, "function

23 isEven(n) {", it interprets the prompt as the beginning of a function written in the JavaScript language that

24 will test whether a number is even, same as Codex.

25   81.   However, the Output of Copilot in response to the prompt is different than Codex, namely:

26 function isEven(n) {

27   return n % 2 === 0;
   }
28

82.   This function is much closer to what a human programmer might use as compared to Codex's older, inaccurate offering. It handles all values and types of "n" correctly. It does not cause a stack overflow for larger values of "n" like the Codex Output.

83.   Copilot's Output, like Codex's, is derived from existing code. Namely, sample code that appears in the online book *Mastering JS*, written by Valeri Karpov.[10] Like Codex's Output, Copilot's is also based upon copyrighted educational material. *Mastering JS* is a set of educational exercises for programmers. Like *Eloquent JavaScript*, there are many copies of Karpov's exercise stored in public repositories on GitHub. Programmers working through *Mastering JS* store their answers there.

84.   If Copilot is prompted with the name of a function that will test whether a number is prime (that is, a number that can only be evenly divided by 1 and itself), namely "function isPrime(n) {", it returns:

```
function isPrime(n) {
  if (n < 2) {
    return false;
  }
  for (let i = 2; i < n; i++) {
    if (n % i === 0) {
      return false;
    }
  }
  return true;
}
```

85.   Though this function will work, it contains an error often made by beginner programmers that makes it much slower than it could be. Namely, the loop in the middle, which checks possible divisors, does not need to check every divisor smaller than "n," only the divisors smaller than the square root of "n". As with Codex, Copilot has no understanding of how the code works. It knows that more functions called "isPrime" contain the portion that checks for all divisors smaller than "n", so that is what it offers. It does not return what it "thinks" is best, it returns what it has seen *the most*. It is not writing, it is reproducing (i.e., copying).

___

[10] https://masteringjs.io/tutorials/fundamentals/modulus/.

86.   Like the other examples above—and most of Copilot's Output—this output is nearly a verbatim copy of copyrighted code. In this case, it is substantially similar to the "isPrime" function in the book *Think JavaScript* by Matthew X. Curinga et al,[11] which is:

```
function isPrime(n) {
   if (n < 2) {
      return false;
   }
   for (let i = 2; i < n; i++) {
      if (n % i === 0) {
         return false;
      }
   }
   return true;
}
```

87.   As with the other examples above, the source of Copilot's Output is a programming textbook. Also like the books the other examples were taken from, there are many copies of Curinga's code stored in public repositories on GitHub where programmers who are working through Curinga's book keep copies of their answers.

88.   The material in Curinga's book is made available under the GNU Free Documentation License. Although this is not one of the Suggested Licenses, it contains similar attribution provisions, namely that "You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License."[12]

89.   As with Codex, Copilot does not provide the end user any attribution of the original author of the code, nor anything about their license requirements. There is no way for the Copilot user to know that they must provide attribution, copyright notice, nor a copy of the license's text. And with regard to the GNU Free Documentation License, Copilot users would not be aware that they are limited in what conditions they can place on the use of derivative works they make using this copyrighted code. Had the

---

[11] https://matt.curinga.com/think-js/#solving-problems-with-for-loops.

[12] https://matt.curinga.com/think-js/#gnu-free-documentation-license.

1 Copilot user found this code in a public GitHub repository or a copy of the book it was originally

2 published in, they would find the GNU Free Documentation License at the same time and be aware of its

3 terms. Copilot finds that code for the user but excises the license terms, copyright notice, and attribution.

4 This practice allows its users to assume that the code can be used without restriction. It cannot.

5 **D. Codex and Copilot Were Trained on Copyrighted Materials Offered Under Licenses**

6 90.   Codex is an AI system. Another way to describe it is a "model." Without Codex, Copilot,

7 or another AI-code-lookup-tool, code is written both by originating code from the writer's own knowledge

8 of how to write code as well as by finding pre-written portions of code that—under the terms of the

9 applicable license—may be incorporated into the coding project.

10 91.   Unlike a human programmer that has learned how code works and notices when code it is

11 copying has attached license terms, a copyright notice, and/or attribution, Codex and Copilot were

12 developed by feeding a corpus of material, called "training data," into them. These AI programs ingest all

13 the data and, through a complex probabilistic process, predict what the most likely solution to a given

14 prompt a user would input is. Though more complicated in practice, essentially Copilot returns the

15 solution it has found in the most projects when those projects are somehow weighted to adjust for

16 whatever variables Codex or Copilot have identified as relevant.

17 92.   Codex and Copilot were not programmed to treat attribution, copyright notices, and

18 license terms as legally essential. Defendants made a deliberate choice to expedite the release of Copilot

19 rather than ensure it would not provide unlawful Output.

20 93.   The words "study" and "training" and "learning" in connection with AI describe

21 algorithmic processes that are not analogous to human reasoning. AI models cannot "learn" as humans

22 do, nor can it "understand" semantics and context the way humans do. Rather, it detects statistically

23 significant patterns in its training data and provides Output derived from its training data when

24 statistically appropriate. A "brute force" approach like this would not be efficient nor even possible for

25 humans. A human could not memorize, statistically analyze, and easily access thousands of gigabytes of

26 existing code, a task now possible for powerful computers like those that make up Microsoft's Azure cloud

27 platform. To accomplish the same task, a human may search for Licensed Materials that serve their

28 purpose if they believe such materials exist. And if that human finds such materials, they will probably

abide by its License Terms rather than risk infringing its owners' rights. At the very least, if they incorporate those Licensed Materials into their own project without following its terms they will be doing so knowingly.

E. Copilot Was Launched Despite Its Propensity for Producing Unlawful Outputs

94.   GitHub and OpenAI have not provided much detail regarding what data Codex and OpenAI were trained on. Plaintiffs know for certain from GitHub and OpenAI's statements, that both systems were trained on publicly available GitHub repositories, with Copilot having been trained on all available public GitHub repositories.

95.   According to OpenAI, Codex was trained on "billions of lines of source code from publicly available sources, including code in public GitHub repositories." Similarly, GitHub has described[13] Copilot's training material as "billions of lines of public code." GitHub researcher Eddie Aftandilian confirmed in a recent podcast[14] that Copilot is "train[ed] on public repos on GitHub."

96.   In a recent customer-support message, GitHub's support department clarified certain facts about training Copilot. First, GitHub said that "training for Codex (the model used by Copilot) is done by OpenAI, not GitHub." Second, in its support message, GitHub put forward a more detailed justification for its use of copyrighted code as training data:

> Training machine learning models on publicly available data is considered fair use across the machine learning community . . . OpenAI's training of Codex is done in accordance with global copyright laws which permit the use of publicly accessible materials for computational analysis and training of machine learning models, and do not require consent of the owner of such materials. Such laws are intended to benefit society by enabling machines to learn and understand using copyrighted works, much as humans have done throughout history, and to ensure public benefit, these rights cannot generally be restricted by owners who have chosen to make their materials publicly accessible.

The claim that training ML models on publicly available code is widely accepted as fair use is not true. And regardless of this concept's level of acceptance in "the machine learning community," under Federal law, it is illegal.

---

[13] https://github.blog/2021-06-30-github-copilot-research-recitation/.

[14] https://www.se-radio.net/2022/10/episode-533-eddie-aftandilian-on-github-copilot/.

97.   Former GitHub CEO Nat Friedman said in June 2021—when Copilot was released to a limited number of customers—that "training ML systems on public data is fair use."[15] Friedman's statement is pure speculation; no Court has considered the question of whether "training ML systems on public data is fair use." The Fair Use affirmative defense is only applicable to Section 501 copyright infringement. It is not a defense to violations of the DMCA, breach of contract, nor any other claim alleged herein. It cannot be used to avoid liability here. At the same time Friedman asserted "the output [of Copilot] belongs to the operator."

98.   Other open-source stakeholders have made this point already. For example, in June 2021, Software Freedom Conservancy ("SFC"), a prominent open-source advocacy organization, asked Microsoft and GitHub to provide "legal references for GitHub's public legal positions." No references were provided by any of the Defendants.[16]

99.   Beyond the examples above, Copilot regularly Output's verbatim copies of Licensed Materials. For example, Copilot reproduced verbatim well-known code from the game Quake III, use of which is governed by one of the Suggested Licenses—GPL-2.[17]

100.   Copilot also reproduced code that had been released under a license that allowed its use only for free games and required attribution by including a copy of the license. Copilot did not mention nor include the underlying license when providing a copy of this code as Output.[18]

101.   Texas A&M computer-science professor Tim Davis has provided numerous examples of Copilot reproducing code belonging to him without its license or attribution.[19]

102.   GitHub concedes that in ordinary use, Copilot will reproduce identical passages of code, verbatim: "Our latest internal research shows that about 1% of the time, a suggestion [Output] may contain some code snippets longer than ~150 characters that matches" code from the training data. This standard is more limited than is necessary for copyright infringement. But even using GitHub's own

---

[15] https://twitter.com/natfriedman/status/1409914420579344385/.

[16] https://sfconservancy.org/blog/2022/feb/03/github-copilot-copyleft-gpl/.

[17] https://twitter.com/stefankarpinski/status/1410971061181681674/.

[18] https://twitter.com/ChrisGr93091552/status/1539731632931803137/.

[19] https://twitter.com/DocSparse/status/1581461734665367554/.

1 metric and the most conservative possible criteria, Copilot has violated the DMCA at least tens of

2 thousands of times.

3   103.   In June 2022, Copilot had 1,200,000 users. If only 1% of users have ever received Output

4 based on Licensed Materials and only once each, Defendants have "only" breached Plaintiffs' and the

5 Class's Licenses 12,000 times. However, each time Copilot outputs Licensed Materials without

6 attribution, the copyright notice, or the License Terms it violates the DMCA three times. Thus, even

7 using this extreme underestimate, Copilot has "only" violated the DMCA 36,000 times. Because Copilot

8 constantly Outputs code as a user writes, and because nearly all of Copilot's training data was Licensed

9 Material, this number is most likely exponentially lower than the true number of breaches and DMCA

10 violations.

11   104.   Academics are continuing to study generative AI models and their behavior. Recent

12 academic research shows that the likelihood Plaintiffs' or class members' code would be emitted verbatim

13 is only increasing. For instance the study, *Quantifying Memorization Across Neural Language Models* by

14 Nicholas Carlini et al.,[20] tested multiple models by feeding prefixes of prompts based on training data into

15 each model in order to compare the performance of models of different sizes to emit output that is

16 identical to training data. The study concluded:

17
18
19
20
21
22
23
> Large language models (LMs) have been shown to memorize parts of their
> training data, and when prompted appropriately, they will emit the
> memorized training data verbatim. … We describe three [mathematical]
> relationships that quantify the degree to which LMs emit memorized
> training data. **Memorization significantly grows as we increase (1)
> the capacity of a model**, (2) the number of times an example has been
> duplicated, and (3) the number of tokens of context used to prompt the
> model. Surprisingly, we find the situation becomes more complicated
> when generalizing these results across model families. On the whole, we
> find that memorization in LMs is more prevalent than previously believed
> and will likely get worse as models continues to scale.

24 (Emphasis added). Or as simply put by the study, "Bigger Models Memorize More."

25

26

27

---

[20] https://arxiv.org/pdf/2202.07646.pdf

28

105.   In other words, as generative AI models such as Copilot increase capacity and continue to scale, it becomes **more likely** that training data will become memorized and emitted verbatim, i.e., as an exact duplicate.

106.   Given GitHub's increasing commitment to Copilot, and the scale of growth of Copilot, it follows that Copilot is more likely to emit duplicates of memorized training data as it continues to scale.

107.   Indeed, GitHub has announced that it is adopting OpenAI's GPT-4 model for Copilot, which is a bigger and more capable language model than the Codex-derived model.

108.   Furthermore, the Suggested Licenses impose attribution obligations not only when Licensed Materials have been used verbatim, but also when Licensed Materials have been modified or adapted. Though Output from Copilot is often a verbatim, i.e., identical copy, even more often it is a modification: for instance, a near-identical copy that contains only semantically insignificant variations of the original Licensed Materials, or a modified copy that recreates the same algorithm. Whenever Copilot outputs Licensed Materials in a manner that qualifies as a modification, the attribution requirements of the Suggested Licenses still apply. Copilot's failure to provide the attributions for outputs that are modifications of Licensed Materials represents another enormous set of license breaches.

F.   Copilot Reproduces the Code of the Named Plaintiffs Without Attribution

109.   Because Copilot was trained on all available public GitHub repositories, if Licensed Materials have been posted to a GitHub public repository, Plaintiffs and the Class can be reasonably certain it was ingested by Copilot and is sometimes returned to users as Output.

110.   Described below are some specific examples of Copilot's unlawful behavior using Licensed Materials owned by the named Plaintiffs. These examples were emitted by Copilot after prompting Copilot.

111.   In the examples below, original code is shaded gray, prompts to Copilot are shaded orange, and outputs from Copilot are shaded light blue.

1.   Example: Copilot Outputs the Code of Doe 2 Essentially Verbatim

112.   The first example demonstrates Copilot suggesting an essentially verbatim copy of code written by Doe 2.

[CONFIDENTIAL]

113. ███ subject to the GNU General Public License v3.0. ████████████████████ ███████████████████████████████████ ████████████████████████████████████ The relevant code from the original source file is shown below:

██████████████████

114.   When Copilot is prompted the first few lines of Doe 2's code:

████████████████

████████████████████████

Copilot suggests the following:

██████████████████

[CONFIDENTIAL]

115.   This suggestion from Copilot is identical to Doe 2's code, except that ████████████

███

███

These differences in the code are cosmetic and the code is functionally equivalent; otherwise, this is a

verbatim copy. Doe 2's particular arrangement and sequencing seen in his code is distinctive expression

found only in one location on GitHub: ████████████████████

116.   Because the Copilot suggestion is a nearly verbatim reproduction of Doe 2's unique code,

it follows that Copilot copied Doe 2's code. Copilot therefore needed to adhere to the requirements of

Doe 2's license (GNU General Public License v3.0) for that code, including providing attribution. It does

not. Copilot also did not reproduce Doe 2's license.

2.   Example: Copilot Outputs the Code of Doe 1 in Modified Format

117.   The second example demonstrates Copilot suggesting a modified copy of code written by

Doe 1. To protect Doe 1's identity, the paragraphs describing the code will be redacted.

118.   ██ subject to the

MIT License. ███

████████████████

[CONFIDENTIAL]

SECOND AMENDED COMPLAINT

[CONFIDENTIAL]

119.   When Copilot is prompted with ███████████████████████████████

The first suggestion from Copilot is a modification of Doe 1's code:

[CONFIDENTIAL]

120.   ██ do not

appear in any other source file on GitHub. The only way Copilot knows how to make this suggestion is

because it ingested Doe 1's source file as training data. Though the Copilot suggestion is not an exact

match for Doe 1's code, it is necessarily a modification based on a copy of Doe 1's code.

121.   Furthermore, many distinctive expressive features of Doe 1's code have been preserved in

Copilot's suggestion. For instance, Doe 1's comments in the code (in green) are reproduced almost

verbatim. ██

██

██

[CONFIDENTIAL]

1

2   means the same thing as this Copilot-suggested code:

3

4

5   122.   As is apparent from a cursory glance of this example, the variations between Copilot's

6   emitted output and Doe 1's source code are cosmetic and the code is functionally equivalent; it follows

7   that Copilot's output is a copy of Doe 1's code.

8   123.   That said, Copilot also introduces mistakes into the code. For instance,

9

10

11   124.   Still, because Copilot is reproducing Doe 1's algorithm in modified format, and the

12   obligations in Doe 1's license (the MIT License) carry with the code even if the underlying code is

13   modified, the Copilot suggestion needs to follow the requirements of Doe 1's license for that code,

14   including providing attribution. It does not. Copilot also did not reproduce Doe 1's license.

            3.   **Example: Copilot Outputs the Code of Doe 5 In Modified Format**

15

16   125.   The third example demonstrates Copilot suggesting multiple modified copies of code

17   written by Doe 5 in response to a sequence of prompts, which is a common way of using Copilot. To

18   protect Doe 5's identity, the paragraphs describing the code will be redacted.

19   126.

20   subject to the MIT License.

21   The relevant code from the original source file is shown below:

22

23

24

25

26

27

28

[CONFIDENTIAL]

127.   When Copilot is prompted the first section of Doe 5's code, comprising the first complete test and the name of the second:

128.   The first suggestion from Copilot offers to complete the prompt with a verbatim copy of Doe 5's original code, except that ███████████████████████████████ (a variation that does not affect how the code works):

129.   Next, if the name of the third test is appended, the next prompt to Copilot looks like this:

[CONFIDENTIAL]

130.   The first suggestion from Copilot offers to complete the prompt with a functionally identical copy of Doe 5's code, except ███████████████████████████████████████ ██████████████████████████ (neither of these variations affect how the code works):

131.   As is apparent from the high degree of similarity and minor cosmetic deviations between Copilot's emitted output and Doe 5's source code, Copilot ingested, copied and reproduced Doe 5's source code as output.

132.   Because Copilot is (repeatedly) reproducing Doe 5's original code in modified format, and the obligations in Doe 5's license (the MIT License) carries with the code even when it is modified, the Copilot suggestions need to follow the requirements of Doe 5's license for that code, including providing attribution. They do not. Copilot also did not reproduce Doe 5's license.

[CONFIDENTIAL]

4.   Example: Copilot Outputs Code of Doe 5 Essentially Verbatim

133.   The fourth example also demonstrates Copilot suggesting multiple modified copies of code written by Doe 5 in response to a sequence of prompts, which is a common way of using Copilot. To protect Doe 5's identity, the paragraphs describing the code will be redacted.

134.   ██

████████████ subject to the MIT License. ████████████████████████████

████████████████ The first three tests from the original source file are shown below:

SECOND AMENDED COMPLAINT

[CONFIDENTIAL]

135.   When Copilot is prompted with the first section of Doe 5's code, comprising the first complete test and the name of the second:

The first suggestion from Copilot offers to complete the second test with a verbatim copy of Doe 5's original code:

SECOND AMENDED COMPLAINT
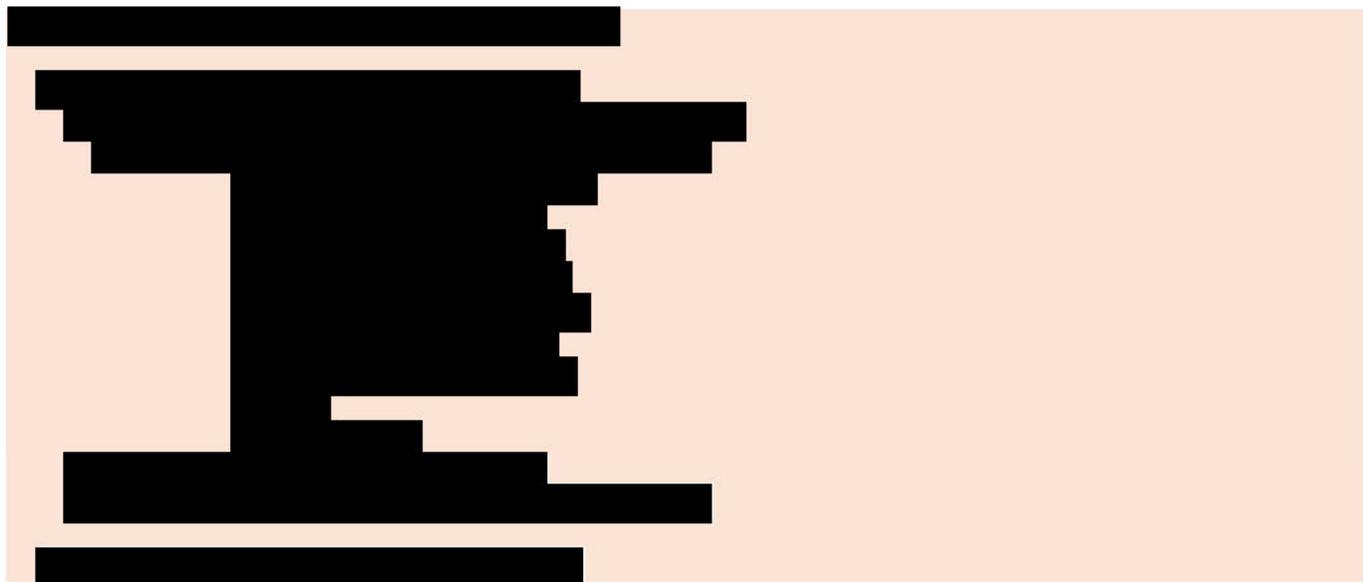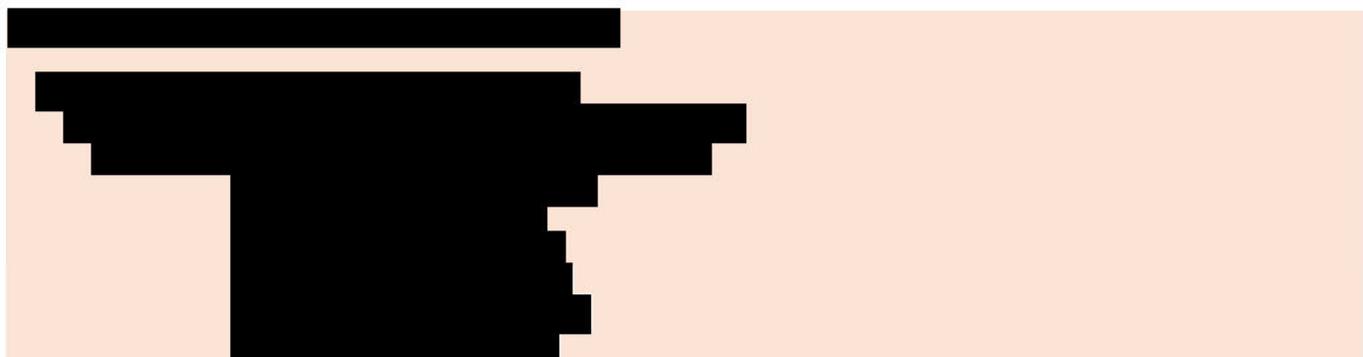
[CONFIDENTIAL]

136.   When Copilot's suggestion is accepted and the name of Doe 5's third test is appended, the next prompt to Copilot looks like this:

SECOND AMENDED COMPLAINT

[CONFIDENTIAL]

137.   Once again, the first suggestion from Copilot offers to complete the third test with a verbatim copy of Doe 5's code (except for small cosmetic variations in line breaks):

[CONFIDENTIAL]

138.   Because Copilot is (repeatedly) reproducing Doe 5's code essentially verbatim, the Copilot suggestions need to follow the requirements of Doe 5's license (the MIT License) for that code, including providing attribution. They do not. Copilot also did not reproduce Doe 5's license.

139.   These are only a few examples of Plaintiffs' code being reproduced by Copilot. It follows that many if not all prompts entered into Copilot will readily cause it to emit verbatim, near-verbatim or modified copies of Licensed Material that violate the licenses under which the source code is published. Multiplied across the many users of Copilot and the many times Copilot is prompted, each day these violations must be accruing with astonishing frequency. It is therefore likely if not certain that verbatim, near-verbatim or modified copies of each Plaintiffs' code have already been emitted by Copilot.

140.   Additionally, even though Plaintiffs have been able to generate these examples, Plaintiffs remain at a great evidentiary disadvantage relative to Defendants, because Defendants control all the information about the training dataset. In particular, only Defendants know *when* the Licensed Materials of Plaintiffs and the Class were scraped. As is typical in open source, many of the Licensed Materials are regularly updated. As such, it is difficult to determine which iterations of code may have been trained on and would be subject to emission by Copilot.

G. Codex and Copilot Were Designed to Withhold Attribution, Copyright Notices, and License Terms from Their Users

141.   Codex and Copilot have no way to determine whether license text or other Copyright Management Information ("CMI")[21] is part of the code it appears immediately before or after. Unless instructed otherwise, it will assume that CMI that usually appears just before a given block of code is an important part of that code or otherwise necessary for it to function.

142.   It is a common practice to provide the applicable license text at the top of every source file in the codebase. The purpose of this practice is to avoid the code from being divorced from the license. This may occur via "vendoring," a method of creating a derivative work by including source files from a

---

[21] CMI is defined in detail below in Paragraph 211.

copyrighted project directly into another project without following the terms of the license or providing attribution or a copyright notice. Copilot circumvents this protective measure to mask the degree of vendoring it engages in.

143.   Early iterations of Copilot reproduced license text. For example, in a blog post, GitHub noted "In one instance, GitHub Copilot suggested starting an empty file with something it had even seen more than a whopping 700,000 different times during training–that was the GNU General Public License."[22] Copilot no longer suggests licenses in this way because it has been altered not to. As GitHub explains: "GitHub Copilot *has* changed to require a minimum file content. So some of the suggestions flagged here would not have been shown by the current version."

144.   In July 2021, near Copilot's launch, it would sometimes produce license text, attribution, and copyright notices. This CMI was not always accurate. Copilot no longer reproduces these types of CMI, incorrect or otherwise, on a regular basis. It has been altered not to.

145.   In July 2022, in response to public criticism of Copilot's mishandling of Licensed Materials, GitHub introduced a user-settable Copilot filter called "Suggestions matching public code." If set to "block," this filter claims to prevent Copilot from suggesting verbatim excerpts of "about 150 characters" that come from Licensed Materials. But even assuming the filter works as advertised, because it only checks for verbatim excerpts, it does nothing to impede the Outputs from Copilot that are modifications of Licensed Materials. Thus, as a device for respecting the rights of Plaintiffs and the Class, it is essentially worthless.

146.   GitHub Copilot now "includes an option to either allow or block code completion suggestions that match publicly available code. If you choose to block suggestions matching public code, GitHub Copilot checks code completion suggestions with their surrounding code of about 150 characters against public code on GitHub. If there is a match, or a near match, the suggestion is not shown to you."[23] There is no reason provided for GitHub's choice of 150 matching characters of code. Nor is there a reason for GitHub to include an option to block suggestions that "match" code unless Copilot is capable of (and

---

[22] https://github.blog/2021-06-30-github-copilot-research-recitation/.

[23] https://docs.github.com/en/copilot/configuring-github-copilot/configuring-github-copilot-settings-on-githubcom#enabling-or-disabling-duplication-detection

1 does) emit verbatim copies of code. Nonetheless, GitHub provides the choice to Copilot's paying users to

2 use code Copilot outputs that is identical to code on its public repositories, subject to open source

3 licenses.

4   147.   GitHub now admits[24] that "You can opt to allow GitHub Copilot to suggest code

5 completions that match publicly available code on GitHub.com." GitHub states that "If you have allowed

6 suggestions that match public code, GitHub Copilot can provide you with details about the matching code

7 when you accept such suggestions. This feature is called code referencing." Again, the only reason

8 GitHub would inform users that they can opt in to allowing Copilot to suggest code completions that

9 "match" publicly available code (i.e., code that Codex and Copilot were trained on) is that Copilot is

10 capable of, and does, emit code suggestions that are verbatim copies of code.

11   148.   GitHub states it now can provide its users, at their option, a link in to the relevant identical

12 open-source license under which such identical code was published on GitHub. Having acknowledged

13 there is the likelihood, if not certainty that Github will produce exact copies, and providing a tool to

14 prevent this, GitHub makes it entirely optional to users, and provides no such optionality to licensors.

15 Thus users who want to receive identical code from GitHub or do not want to exclude it, may do so. In so

16 doing, GitHub facilitates and encourages users to receive identical code.

17   149.   As a result, Plaintiffs are informed and believe, and on that basis allege that it is likely that

18 their licensed code is omitted by Github in violation of the open source licenses. Further Plaintiffs are

19 informed and believe, and on that basis allege that with respect to numerous members of the class, it is

20 certain that some identical code has been omitted by GitbHub. Plaintiffs are informed and believe, and on

21 that basis allege that there is a substantial risk, if not certainty, that identical code will be emitted in the

22 future. Further, given the fact that GitHub has implemented a tool to prevent this, and have made it

23 optional, not mandatory, to users, GitHub knows, or has reason to know that identical code will be

24 omitted in the future.

25   150.   Further, Github but makes clear it is entirely up the user to add any license or attribution to

26 the code GitHub generated for them. "The linked web page includes details of any license identified for

27 ─────────────────────
[24] https://docs.github.com/en/copilot/using-github-copilot/finding-public-code-that-matches-github-

28 copilot-suggestions

1   the repository where the matching code was found. Having reviewed the references, *you can decide how to*

2   *proceed.*" This link temporarily shows up in the users' "GitHub Copilot Log view" but "[t]he GitHub

3   Copilot log is flushed when you close the editor."

4   151.   In addition to the fact that the tool which identifies and screens identical code is an option,

5   at *the discretion of the user*, GitHub also acknowledges that the optional tool to prevent or exclude identical

6   code is not 100 per cent effective. Indeed, GitHub confirms it is limited in its scope and might produce

7   Plaintiffs or class members' matching code, but otherwise will not detect it under certain circumstances.

8   GitHub states "code from public repositories deleted before the index was created, may not be included in

9   the search. For the same reason, the search may return matches to code that has been deleted or moved

10   since the index was created."

11   152.   GitHub also admits the ability and therefore the efficacy of the code referencing tool is

12   limited, incomplete and predictably ineffective, admits that its code referencing feature is limited to

13   finding identical public code made from indexes of GitHub public repositories *after* training Copilot, such

14   that "code from public repositories deleted before the index was created, may not be included in the

15   search." Github thereby admits that Copilot trained on code posted on GitHub can be output in identical

16   form, without the required attribution of other compliance with license terms. This may occur whch

17   nonetheless still may not be attributable in any way, even with the best intentions of Copilot's commercial

18   users and even given GitHub's "code reference" feature . Simply put, GitHub admits Copilot can and

19   does output identical matching code of Plaintiffs and class members that its own "code referencing"

20   feature cannot detect, even on the infrequent occasion when a user exercises the option to implement the

21   feature.

22   153.   GitHub further represents to Copilot's paying users that "Typically, matches to public

23   code occur in less than one percent of Copilot suggestions, so you should not expect to see code

24   references for many of the suggestions you accept." On information and belief, Plaintiffs are informed abd

25   believe, and allege that this number is much higher.

26   154.   GitHub elsewhere represents: "If you choose to allow suggestions matching public code,

27   and you accept a suggestion for which one or more matches were found, you can click through from an

28

1 entry in the GitHub Copilot log to view a list of references on GitHub." [25] But there is no requirement to

2 attach the license to matching code Copilot outputs for its paying users. In other words, there are many

3 situations in which public code is emitted without compliance with license terms.

4   155. In GitHub's hands, the propensity for small cosmetic variations in Copilot's Output is a

5 feature, not a bug. GitHub does so knowing that these small cosmetic variations allow CoPilot to conceal

6 the copying of Licensed Materials and to separate the Licensed Material from the licenses. In so doing,

7 GitHub knowingly conceals the copying of Licensed Materials with the intent and purpose of making it

8 difficult for Plaintiffs and members of the class to identify breaches of the licenses and to enforce their

9 rights. These small cosmetic variations mean that GitHub can deliver to Copilot customers unlimited

10 modified copies of Licensed Materials without ever triggering Copilot's verbatim-code filter. AI models

11 like Copilot often have a setting called *temperature* that specifically controls the propensity for variation in

12 their output. On information and belief, GitHub has optimized the temperature setting of Copilot to

13 produce small cosmetic variations of the Licensed Materials as often as possible, so that GitHub can

14 deliver code to Copilot users that *works* the same way as verbatim code, while claiming that Copilot only

15 produces verbatim code 1% of the time. This technique of active concealment has been effective on many

16 occasions. Copilot is an ingenious method of software piracy and concealment.

17   156. In December 2022, GitHub launched Copilot for Business. The initial terms of service

18 included one notable extra provision compared to ordinary Copilot: a "Defense of Third Party Claims"

19 that read:

20   GitHub will defend you against any claim by an unaffiliated third-party that

21   your use of GitHub Copilot misappropriated a trade secret or directly
infringes a patent, copyright, trademark, or other intellectual property right

22   of a third party, up to the greater of $500,000.00 USD or the total amount
paid to GitHub for the use of GitHub Copilot during the 12 months

23   preceding the claim. GitHub's defense obligations do not apply if (i) the
claim is based on Code that differs from a Suggestion provided by GitHub

24   Copilot, (ii) you fail to follow reasonable software development review
practices designed to prevent the intentional or inadvertent use of Code in a

25   way that may violate the intellectual property or other rights of a third party,
or (iii) you have not enabled all filtering features available in GitHub

26   Copilot.

27 ───────────────────────────
[25] https://docs.github.com/en/copilot/configuring-github-copilot/configuring-github-copilot-settings-on-

28 githubcom#enabling-or-disabling-duplication-detection

157.   If Copilot had been designed to reproduce the attribution, license terms, and copyright notices of the Licensed Materials, this kind of contractual reassurance wouldn't be necessary. With this provision (since removed), GitHub acknowledged that Copilot disrupts—possibly with legal consequences—the relationship between authors and users of open-source software.

B.   Open-Source Licenses Began to Appear in the Early 1990s

158.   In 1991, software engineer Linus Torvalds began a project to create a UNIX-like operating system that would run on common PC hardware. This project became known as Linux.

159.   To encourage adoption of his system, and persuade other programmers to contribute, he released Linux under what was then an unusual software license called the GNU General Public License, or GPL.

160.   The GPL is a software license. But whereas most software licenses required payment, software under the GPL is provided for free. Whereas most software licenses did not include source code, GPL software always included source code. And whereas most software licenses prohibited derivative works, the GPL not only allowed it, but encouraged it.

161.   In certain ways, however, the GPL still operated like a traditional software license. For example, consistent with copyright law, it depended on an assertion of copyright by the software author. Even though GPL software was available at no charge, the GPL contained conditions on its users as licensees.

162.   One license requirement was that a program derived from GPL software had to redistribute certain information about that software:

> You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this General Public License and to the absence of any warranty; and give any other recipients of the Program a copy of this General Public License along with the Program.[26]

Failure to adhere to these conditions constituted a violation of the license, triggering the possibility of

---

[26] https://www.gnu.org/licenses/old-licenses/gpl-1.0.en.html.

1   legal action. Provisions of the GPL are enforceable, and many GPL licensors have sought to enforce GPL

2   licenses though court proceedings and other litigation.

3   163.   The early years of Linux paralleled the early years of the World Wide Web. The fact that

4   Linux was free and ran on common computer hardware made it a popular choice for web servers. Because

5   of its contrarian GPL licensing, Linux became hugely popular. A large ecosystem of other programs and

6   tools grew around it. This contributed to the explosive growth of the web and other network services

7   across the rest of the 1990s.

8   164.   In turn, the growth of the World Wide Web made it easier for developers in different places

9   to collaborate on software. The GPL, and licenses like it, were a natural fit for this kind of collaborative

10   work.

11   165.   Around 1998, a new name was coined as an umbrella term for these principles of software

12   licensing and development: *open source*.

13   **H. Microsoft Has a History of Flouting Open-Source License Requirements**

14   166.   During the 1980s and 1990s, Microsoft was primarily a software company, focusing largely

15   on operating systems and related applications. These included its DOS operating system and later, its

16   Windows operating system. Windows generated billions of dollars in revenue from its sale and licensing as

17   proprietary software for desktop computers and servers. Microsoft derived substantial income from sale

18   of licensed products and devotes substantial resources to protecting and enforcing such licenses.

19   167.   Windows is a graphical operating system. It allows users to view and store files, run

20   software and games, play videos, and provides a way to connect to the internet.

21   168.   Linux represented a competitive threat to Windows. It ran on the same hardware. It

22   performed many of the same functions. It was free. Many programmers at the time considered Linux to be

23   functionally superior to Windows.

24   169.   Microsoft has engaged in a problematic practice known as "vaporware," where products

25   are announced but are in fact late, never manufactured, or canceled. Typically the company promising

26   vaporware never has any intention of providing it. The term vaporware was coined by Microsoft in 1982 in

27   reference to the development of its Xenix operating system.

28

170.   Microsoft described its anti-Linux strategy as "FUD," standing for fear, uncertainty, and doubt. Microsoft focused extra attention to Linux's open-source aspects.

171.   In 1998, a source at Microsoft leaked what became known as the "Halloween Documents", revealing Microsoft's thinking on how to counter the competitive threat from Linux. Among other things, the documents emphasized the importance of countering the "long term developer mindshare threat", and concluded that to defeat open source, "[Microsoft] must target a process rather than a company."[27]

172.   In 2001, Microsoft CEO Steve Ballmer said "The way the [GPL] is written, if you use any open-source software, you must make the rest of your software open source. . . . Linux is a cancer that attaches itself in an intellectual property sense to everything it touches."[28] Ballmer's summary of GPL licensing was not accurate. In 2001, Linux was being used by corporations of every size. The growth of open source up to that point, and since, has been made possible by the open-source community's respect for and compliance with applicable licenses.

173.   In 2001, Microsoft was the defendant in a major software-related antitrust case, *United States v. Microsoft Corporation*.[29] In this case, the U.S. Department of Justice accused Microsoft of maintaining a software monopoly by illegally imposing technical restrictions on manufacturers of personal computers, including "tying" violations related to the Internet Explorer web browser. Judge Thomas Penfield Jackson, who presided over the antitrust trial, opined that Microsoft is "a company with an institutional disdain for both the truth and for rules of law that lesser entities must respect. It is also a company whose 'senior management' is not averse to offering specious testimony to support spurious defenses to claims of its wrongdoing."[30]

174.   In 2007, Microsoft admitted that it tried to influence the vote of an ISO open-standards committee by offering money to certain business partners in Sweden to vote for Microsoft's preferred outcome.[31]

---

[27] http://www.catb.org/esr/halloween/halloween1.html.

[28] https://lwn.net/2001/0607/a/esr-big-lie.php3.

[29] No. Civ.A. 00–1457 TPJ.

[30] *Jackson v. Microsoft Corp.*, 135 F. Supp. 2d 38 (D.D.C. 2001).

[31] https://learn.microsoft.com/en-us/archive/blogs/jasonmatusow/open-xml-the-vote-in-sweden/.

175.   After observing the rapid growth of Amazon's original cloud computing products, Microsoft has expanded its business into cloud computing, which it has branded Microsoft Azure or simply Azure. Microsoft announced Azure to developers in 2008. It was formally released in 2010. Azure uses large-scale virtualization at Microsoft data centers and offers many hundreds of services, including infrastructure as a service ("IaaS"), platform as a service ("PaaS"), compute services, Azure Active Directory, mobile services, storage services, communication services, data management, messaging, developer services, Azure AI, blockchain, and others.

I.   GitHub Was Designed to Cater to Open-Source Projects

176.   By 2002, Linux had become immensely popular. But the project itself had become unwieldy and had outgrown its reliance on informal systems of managing software source code (also known as *source-control systems*). The Linux community needed something better.

177.   Linus Torvalds set about writing a new source-control system. He named his new system Git. He released it under the GPL. It quickly became the source-control system of choice for open-source programmers.

178.   A single software project stored in Git is called a *source repository*, commonly shortened to *repository* or just *repo*. A Git source repository would typically be stored on a networked server accessible to a group of programmers.

179.   This became less convenient, however, when programmers were distributed among multiple locations, rather than being in a single location. A Git repository could be stored on an internet-accessible server. But setting up that server hardware and being responsible for it was inconvenient and expensive.

180.   In 2008, a group of open-source developers in San Francisco, California founded GitHub. GitHub managed internet servers that hosted Git source repositories. With an account at GitHub, an open-source developer could easily set up a Git project accessible to collaborators anywhere in the world. From early on, GitHub's core market was open-source developers, whom it attracted by making many of its hosting services free.

181.   Most open-source programmers used GitHub to create "public" repositories, meaning that anyone could view them & access them. GitHub also allowed programmers and organizations to

1   create "private" repositories, which were not accessible from the public GitHub website, and required

2   password access.

3   182.   Open-source licensing was integral to GitHub. GitHub encouraged open-source developers

4   to understand and use open-source licenses for their work. Many—though not all—public repositories on

5   GitHub carry an open-source license. By convention, this license is stored at the top level of each

6   repository in a file called LICENSE. GitHub's interface also includes a button on the front pages of most

7   repositories users can click to see details of the applicable license. A human user could easily find the

8   license in either of these locations—as could an AI anywhere near as powerful as Codex or Copilot.

9   183.   Though the GPL is one of the early open-source licenses and remains common, it is not

10   the only open-source license. Examples of other common open-source licenses include the MIT License,

11   the Apache License, and the Berkeley Software Distribution License (all of which are included in the

12   Suggested Licenses).

13   184.   Though these licenses differ in their wording and their details, most of them share a

14   requirement that a copy of the license be included with any copy, derivative, or redistribution of the

15   software, and that the author's name and copyright notice remains intact. This is not a controversial

16   requirement of open-source licenses—indeed, it has been an integral part of the GPL for over 30 years.

17   185.   There are also many public repositories on GitHub that have no license. Though GitHub

18   has encouraged awareness of licenses among its users, it has never imposed a default license on public

19   repositories. A public repository without a license is subject to ordinary rules of U.S. copyright.

20   186.   Open-source developers flocked to GitHub. By 2018, GitHub had become the largest and

21   most successful Git hosting service, hosting millions of users and projects.

22   187.   In October 2018, Microsoft acquired GitHub for $7.5 billion. It was important to Microsoft

23   that programmers use GitHub. Microsoft had developed a well-deserved poor reputation because of its

24   documented vaporware, FUD, and other business practices, including those targeted at open-source

25   programs and programming, and open-source licensing specifically. Microsoft made false and misleading

26   statements and omissions to assuage such concerns, including its primary mantra intended to win over the

27   open-source community: "Microsoft Loves Open Source."

28

J.   OpenAI Is Intertwined with Microsoft and GitHub

188.   OpenAI, Inc. is a nonprofit corporation founded in December 2015 by a group that included Greg Brockman, Ilya Sutskever, and other AI researchers; Elon Musk, CEO of Tesla; and Sam Altman, president of Y Combinator, a tech-startup incubator with hundreds of companies in its portfolio. Musk and Altman served as co-chairs of OpenAI, Inc. One of OpenAI, Inc.'s current board members is Reid Hoffman, founder of LinkedIn, which is now a Microsoft subsidiary. Mr. Hoffman is also a member of the Microsoft Board of Directors.

189.   Less than a year later, in November 2016, OpenAI first partnered with Microsoft. It described the partnership as follows: "We're working with Microsoft to start running most of our large-scale experiments on Azure. This will make Azure the primary cloud platform that OpenAI is using for deep learning and AI, and will let us conduct more research and share the results with the world."

190.   Initially, OpenAI, Inc. held itself out as a "non-profit artificial intelligence research company" that sought to shape AI "in the way that is most likely to benefit humanity as a whole."

191.   OpenAI, Inc. reportedly secured $1 billion in initial funding, from sources that were largely not disclosed, but included at least most of its founders.

192.   OpenAI, Inc. obtained its initial source of training data from its founders' companies. According to reporting at the time, Musk and Altman planned to "pool[] online data from their respective companies" to serve as training data for OpenAI, Inc. projects. Musk planned to contribute data from Tesla; Altman planned to have Y Combinator companies "share their data with OpenAI."[32]

193.   In February 2019, Altman created OpenAI, LP, a for-profit subsidiary of the nonprofit entity OpenAI, Inc. The new OpenAI, LP entity would serve as a vessel for accepting traditional outside investment in exchange for equity and distributing profits.

194.   In July 2019, OpenAI, L.P. accepted a $1 billion investment from Microsoft. In addition to cash, Microsoft would become the exclusive licensor of certain OpenAI, LP products (including Codex, described below in Paragraph 197). Also, as part of this alliance, OpenAI, LP would use Microsoft's cloud-computing platform, Azure, exclusively to develop and host its products. Some portion of Microsoft's

---

[32] https://www.wired.com/2015/12/elon-musks-billion-dollar-ai-plan-is-about-far-more-than-saving-the-world/.

investment was paid in credits for use of Azure rather than cash. Finally, Microsoft and OpenAI agreed to "jointly build new Azure AI supercomputing technologies."

195.   Azure is a major growth area for Microsoft. In its most recent earnings report on October 25, 2022, "Azure and other cloud services" grew by 35% from the previous quarter, more than any other product.[33] Azure has grown rapidly since Microsoft began its partnership with OpenAI in 2016. Its revenue grew by 50% or more every quarter from 2016 through the first three quarters of 2020.

196.   In May 2020, Microsoft and OpenAI announced they had jointly built a supercomputer in Azure that would be used exclusively by OpenAI to train its AI models. Microsoft's influence over and frequent collaboration with OpenAI has led some to describe Microsoft as "the unofficial owner of OpenAI."[34]

197.   One of OpenAI's projects is GPT-3, a so-called "large language model" designed to emit naturalistic text. When researchers noticed that GPT-3 could also generate software code, they started studying whether they could make a new AI model specifically trained for this purpose. This project became known as Codex.

198.   Sometime after July 2019, OpenAI and Microsoft began collaborating on a code-completion product for GitHub that would use Codex as its underlying model. This product became known as Copilot.

199.   On September 28, 2022, OpenAI released an image-generation AI called DALL-E-2. Much like Copilot, DALL-E-2 removes any attribution and/or copyright notice from the images it uses to create derivative works. Like with Codex, here, OpenAI ignores the rights of the owners of copyrights to images it has ingested.

200.   In another joint project, Microsoft and OpenAI recently launched a preview of a product called "Azure OpenAI Service."[35] This service will "Leverage large-scale, generative AI models with deep understandings of language and code to enable new reasoning and comprehension capabilities for building

---

[33] https://www.microsoft.com/en-us/Investor/earnings/FY-2023-Q1/press-release-webcast/.

[34] https://venturebeat.com/ai/what-to-expect-from-openais-codex-api/.

[35] https://azure.microsoft.com/en-us/products/cognitive-services/openai-service/.

cutting-edge applications. Apply these coding and language models to a variety of use cases, such as writing assistance, code generation, and reasoning over data. Detect and mitigate harmful use with built-in responsible AI and access enterprise-grade Azure security."

K. Conclusion of Factual Allegations

201.   Future AI products may represent a bold and innovative step forward. GitHub Copilot and OpenAI Codex, however, do not. Defendants should not have released these products until they could ensure that they did not constantly violate Plaintiffs' and the Class's intellectual-property rights, licenses, and other rights.

202.   Defendants have made no attempt to comply with the open-source licenses that are attached to much of their training data. Instead, they have pretended those licenses do not exist, and trained Codex and Copilot to do the same. By simultaneously violating the open-source licenses of tens-of-thousands—possibly millions—of software developers, Defendants have accomplished software piracy on an unprecedented scale. As Microsoft's Co-Founder Bill Gates once said regarding software piracy: "the thing you do is theft."[36]

203.   There is no inherent limitation or constraint of AI systems that made any of this necessary. Defendants chose to build AI systems designed to enhance their own profit at the expense of a global open-source community that they had once sought to foster and protect. GitHub and OpenAI are profiting at the expense of Plaintiffs' and the Class's rights.

VIII.   CLAIMS FOR RELIEF

COUNT 1
VIOLATION OF THE DIGITAL MILLENNIUM COPYRIGHT ACT
17 U.S.C. §§ 1201–1205
(For Injunctive Relief)
(Against All Defendants)

204.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

205.   As described herein, Defendants have intentionally removed or altered CMI from Plaintiffs' code in violation of Section 1202(b)(1) of the DMCA.

---

[36] https://www.digibarn.com/collections/newsletters/homebrew/V2_01/gatesletter.html

206.   As described herein, there is a substantial risk that Defendants will distribute copies of Plaintiffs' code knowing that CMI has been removed or altered while knowing or having reasonable grounds to know that it will induce, enable, facilitate, or conceal infringement in violation of Section 1202(b)(3) of the DMCA.

207.   GitHub has admitted that about 1% of the time, a suggestion may contain code snippets longer than ~150 characters that matches code from the training data. In other words, GitHub itself has admitted that Copilot can emit identical copies of code Copilot was trained on.

208.   GitHub has implemented features which allow users to block output blocks suggestions matching public code. GitHub would not implement such a feature unless it knows that Copilot is capable of, and does, emit output that matches code found on public repositories.

209.   Further, given GitHub's increasing commitment to growing an AI model and the scale of Copilot's code, given that academic research suggests that a model increases the likelihood of emitting training data it has "memorized," the chance that Copilot will emit code that matches code found in the training data is only increasing as the model scales. On information and belief, if Copilot has not done so already, Copilot will emit identical copies of Class members' code.

210.   Plaintiffs and members of the Class own the copyrights to Licensed Materials used to train Codex and Copilot. Copilot was trained on millions—possibly billions—of lines of code publicly available on GitHub. Copilot runs on Microsoft's Azure cloud platform exclusively and Microsoft had input in the creation of Copilot. Microsoft is aware that Copilot ignores License Terms and that it was trained almost exclusively on Licensed Materials.

211.   Plaintiffs and members of the Class included the following Copyright Management Information (as defined in Section 1202(c) of the DMCA) ("CMI") in the Licensed Materials:

    a.   copyright notices;

    b.   the title and other information identifying the Licensed Materials;

    c.   the name of, and other identifying information about, the authors of the Licensed Materials;

    d.   the name of, and other identifying information about, the copyright owners of the Licensed Materials;

e.   terms and conditions for use of the Licensed Materials, specifically the Suggested

Licenses; and

f.   identifying numbers or symbols referring to CMI or links to CMI.

212.   Defendants did not contact Plaintiffs and the Class to obtain authority to remove or alter CMI from the Licensed Materials within the meaning of the DMCA.

213.   Defendants knew that they did not contact Plaintiffs and the Class to obtain authority to remove or alter CMI from the Licensed Materials within the meaning of the DMCA.

214.   As part of the scheme, Defendants did not attempt to contact Plaintiffs or Class members to obtain authority to remove or alter CMI from the Licensed Materials within the meaning of the DMCA. In fact, he removal of CMI made it difficult or impossible to contact Plaintiffs and the Class to obtain authority to remove or alter CMI from the Licensed Materials within the meaning of the DMCA. Rather, Defendants removed or altered CMI from open-source code that is owned by Class members after the code was uploaded to a GitHub repository by incorporating it into Copilot with its CMI removed.

215.   Without the authority of Plaintiffs and the Class, Defendants intentionally removed or altered CMI from the Licensed Materials after they were uploaded to one or more GitHub repositories.

216.   Defendants had access to but were not licensed by Plaintiffs nor the Class to train any machine learning, AI, or other pseudo-intelligent computer program, algorithm, or other functional prediction engine using the Licensed Materials.

217.   Defendants had access to but were not licensed by Plaintiffs nor the Class to incorporate the Licensed Materials into Copilot.

218.   Defendants had access to but were not licensed by Plaintiffs nor the Class to distribute the Licensed Materials as they do through Copilot.

219.   Without the authority of Plaintiffs and the Class, Defendants distributed CMI knowing that the CMI had been removed or altered without authority of the copyright owner or the law with respect to the Licensed Materials.

220.   Defendants distributed copies of the Licensed Materials knowing and intending that CMI had been removed or altered without authority of the copyright owner or the law, with respect to the Licensed Materials.

1   221.   Defendants removed or altered CMI from the Licensed Materials knowing and intending

2   that it would induce, enable, facilitate, or conceal infringement of copyright.

3   222.   Without the CMI associated with the Licensed Materials, Copilot users are induced or

4   enabled to copy the Licensed Materials. Because CMI has been removed, Copilot users do not know

5   whether Output is owned by someone else and subject to restrictions on use. Without the CMI, copyright

6   infringement is facilitated or concealed, because Plaintiffs and the Class are prevented from knowing or

7   learning that the Output is based upon one or more of the Licensed Materials. Use of the Licensed

8   Materials is not infringement when the terms of the applicable Suggested License are followed. Had the

9   CMI not been removed, Copilot users would be aware of the Licenses and their obligations under them.

10   The terms of the applicable Suggested License would have allowed those users to use the Licensed

11   Materials without infringement. By withholding and concealing license information and other CMI,

12   Defendants prevented Copilot users from making non-infringing use of the Licensed Materials. This

13   contradicts the express wishes of Plaintiffs and the Class, which are set forth explicitly in the Suggested

14   Licenses under which the Licensed Materials are offered.

15   223.   Defendants removed or altered CMI from Licensed Materials owned by Plaintiffs and the

16   Class while possessing reasonable grounds to know that it would induce, enable, facilitate, and/or conceal

17   infringement of copyright in violation of Sections 1202(b)(1) and 1202(b)(3) of the DMCA.

18   224.   By omitting, altering and/or concealing CMI from Copilot's Output, Defendants have

19   reasonable grounds to know that innocent infringers are induced or enabled to copy the Licensed

20   Materials, because CMI has been removed. Without the CMI, Defendants have reasonable grounds to

21   know copyright infringement is facilitated or concealed, because Plaintiffs and the Class have the difficult

22   or impossible task of proving the Licensed Materials belong to them.

23   225.   The profits attributable to Defendants' violation of the DMCA include the revenue from:

24   Copilot subscription fees, sales of or subscriptions to Defendants' Copilot-related products and/or

25   services that are used to run Copilot, hosting Copilot on Azure, and any other of Defendants' products

26   that contain copies of the Licensed Materials without all the original CMI. The Licensed Materials add

27   nearly all value to the Copilot product because the purpose of Copilot is to provide code and the source of

28   that code is the Licensed Materials. Without the Licensed Materials, Copilot would not be functional.

226.   On information and belief, Defendants could have trained Copilot to include attribution, copyright notices, and license terms when it provides Output covered by a License.

227.   Defendants did not request or obtain permission from Plaintiffs and the Class to use the Licensed Materials for Defendants' Copilot product.

228.   Defendants use of the Licensed Materials does not follow the requirements of the Suggested Licenses associated with the Licensed Materials. In particular, Copilot fails to provide attribution for the creator nor the owner of the Work. Copilot fails to include the required copyright notice included in the License. Copilot fails to include the applicable Suggested License's text.

229.   Defendants are sophisticated with respect to intellectual property matters related to open-source code. Microsoft in particular has extensive experience granting licenses, obtaining licenses, and enforcing license terms. Its most recent Annual Report states:

> **We protect our intellectual property investments in a variety of ways. We work actively in the U.S. and internationally to ensure the enforcement of copyright, trademark, trade secret, and other protections that apply to our software and hardware products, services, business plans, and branding.** We are a leader among technology companies in pursuing patents and currently have a portfolio of over 69,000 U.S. and international patents issued and over 19,000 pending worldwide. While we employ much of our internally-developed intellectual property exclusively in our products and services, we also engage in outbound licensing of specific patented technologies that are incorporated into licensees' products. From time to time, we enter into broader cross-license agreements with other technology companies covering entire groups of patents. We may also purchase or license technology that we incorporate into our products and services. At times, we make select intellectual property broadly available at no or low cost to achieve a strategic objective, such as promoting industry standards, advancing interoperability, supporting societal and/or environmental efforts, or attracting and enabling our external development community. **Our increasing engagement with open source software will also cause us to license our intellectual property rights broadly in certain situations.**

Microsoft Corporation Annual Report, Form 10-K at 27 (July 28, 2022) (emphasis added).[37]

230.   GitHub, which offers the Copilot product jointly with OpenAI, also has extensive experience with the DMCA. GitHub knows or reasonably should know that the Licensed Materials it hosts are subject to copyright. It provides the language of the Suggested Licenses to users, all of which

---

[37] https://microsoft.gcs-web.com/static-files/07cf3c30-cfc3-4567-b20f-f4b0f0bd5087/.

include copyright notices. Its 2022 Transparency Report—January to June[38] states: "Copyright-related takedowns (which we often refer to as DMCA takedowns) are particularly relevant to GitHub because so much of our users' content is software code and can be eligible for copyright protection."[39] In the first six months of 2022, GitHub processed 1220 DMCA takedown requests. Its DMCA Takedown Policy[40] notes "GitHub probably never would have existed without the DMCA."

231.   GitHub also knows or reasonably should know the portions of the DMCA giving rise to Plaintiffs' claim. In its 2021 Transparency Report, "Before removing content based on alleged circumvention of copyright controls (under Section 1201 of the US DMCA or similar laws in other countries), we carefully review both the legal and technical claims, and we sponsor a Developer Defense Fund to provide developers with meaningful access to legal resources."[41]

232.   GitHub is aware that Copilot's removal of CMI is illegal. For example, it states that "publishing or sharing tools that enable circumvention are not [permitted]"[42] and "Distributing tools that enable circumvention is prohibited, even if their use by developers falls under the exemption [for security research]."[43] GitHub has also frequently published articles discussing the DMCA, its application, and the Copyright Office's guidance on its scope and exceptions.[44]

233.   Unless Defendants are enjoined from violating the DMCA, Plaintiffs and the Class will suffer great and irreparable harm by depriving them of the right to identify and control the reproduction and/or distribution of their copyrighted works, to have the terms of their open-source licenses followed, and to pursue copyright-infringement remedies. Defendants will not be damaged if they are required to comply with the DMCA. Plaintiffs and the Class are therefore entitled to an injunction barring

---

[38] https://github.blog/2022-08-16-2022-transparency-report-january-to-june/.

[39] https://github.blog/2022-08-16-2022-transparency-report-january-to-june/.

[40] https://docs.github.com/en/site-policy/content-removal-policies/dmca-takedown-policy#what-is-the-dmca/.

[41] https://github.blog/2022-01-27-2021-transparency-report/.

[42] https://github.blog/2020-11-19-take-action-dmca-anti-circumvention-and-developer-innovation/#what-dmca-exemptions-do-not-do/.

[43] https://github.blog/2021-11-23-copyright-office-expands-security-research-rights/.

[44] *See, e.g.*, Footnotes 43–46.

Defendants from violating the DMCA and impounding any device or product that is in the custody or control of Defendants and that the court has reasonable cause to believe was involved in a violation of the DMCA.

234.   Defendants conspired together and acted jointly and in concert pursuant to their scheme to commit the acts that violated the DMCA alleged herein.

235.   Defendants induced (or will induce) Copilot users to unknowingly violate the DMCA by withholding attribution, licensing, and other information as described herein.

COUNT 2
BREACH OF CONTRACT—OPEN-SOURCE LICENSE VIOLATIONS
California Common Law
(Against All Defendants)

236.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

237.   Plaintiffs and the Class offer code under various Licenses, the most common of which are set forth in Appendix A. Use of each of the Licensed Materials is allowed only pursuant to the terms of the applicable Suggested License.

238.   Plaintiffs and the Class granted Defendants a license to copy, distribute, and/or create Derivative Works under the Suggested Licenses. Each of the Suggested Licenses requires at least (1) that attribution be given to the owner of the Licensed Materials used, (2) inclusion of a copyright notice for the Licensed Materials used, and (3) inclusion of the terms of the applicable Suggested License. When providing Output, Copilot does not comply with any of these terms.

239.   Defendants accepted the terms of Plaintiffs' and the Class's Licenses when it used the licensed code to create Copilot and when it incorporated the licensed code into Copilot. They have accepted and continue to accept the applicable Licenses every time Copilot Output's Plaintiffs' or the Class's copyrighted code. As such, contracts have been formed between Defendants on the one hand and Plaintiffs and the Class on the other.

240.   Plaintiffs and the Class have performed each of the conditions, covenants, and obligations imposed on them by the terms of the License associated with their Licensed Materials.

1   241.   Plaintiffs and members of the Class hold the copyright in the contents of one or more code

2   repositories that have been hosted on GitHub's platform.

3   242.   Plaintiffs and the Class have appended one of the Suggested Licenses to each of the

4   Licensed Materials.

5   243.   Plaintiffs and the Class did not know about, authorize, approve, or license the Defendants'

6   use of the Licensed Materials in the matter at issue in this Complaint before they were used by

7   Defendants.

8   244.   Defendants have substantially and materially breached the applicable Licenses by failing to

9   provide the source code of Copilot nor a written offer to provide the source code upon the request of each

10   licensee.

11   245.   Defendants have substantially and materially breached the applicable Licenses by failing to

12   provide attribution to the creator and/or owner of the Licensed Materials.

13   246.   Defendants have substantially and materially breached the applicable Licenses by failing to

14   include copyright notices when Copilot Outputs copyrighted OS code.

15   247.   Defendants have substantially and materially breached the applicable Licenses by failing to

16   identify the License applicable to the Work and/or including its text when Copilot Outputs code including

17   a portion of a Work.

18   248.   Plaintiffs and the Class have suffered monetary damages as a result of Defendants'

19   conduct.

20   249.   The conduct of Defendants is causing and, unless enjoined and restrained by this Court,

21   will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully be

22   compensated or measured in money.

23   250.   As a direct and proximate result of these material breaches by Defendants, Plaintiffs and

24   the Class are entitled to an injunction requiring Defendants to comply with all the terms of any License

25   governing use of code that was used to train Copilot, otherwise incorporated into Copilot, and/or

26   reproduced as Output by Copilot.

27   251.   Plaintiffs and the Class are further entitled to recover from Defendants the damages

28   Plaintiffs and the Class sustained—including consequential damages—for Plaintiffs' and the Class's costs

in enforcing their contractual rights. Plaintiffs and the Class are also entitled to recover as restitution from Defendants for any unjust enrichment, including gains, profits, and advantages that Defendants have obtained as a result of their breach of contract.

COUNT 3
BREACH OF CONTRACT — SELLING LICENSED MATERIALS
IN VIOLATION OF GITHUB'S POLICIES
California Common Law
(Against GitHub)

252.   Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

253.   GitHub's Privacy Statement, Terms of Service, and GitHub Copilot Terms share definitions and refer to each other. As such, they are collectively referred to herein as "GitHub's Policies" unless a distinction is necessary and are attached as Exhibit 1.

254.   Plaintiffs and the Class are GitHub users who have accepted GitHub's Policies. As a result, Plaintiffs and the Class have formed a contract with GitHub.

255.   Plaintiffs and the Class have performed each of the conditions, covenants, and obligations imposed on them by the terms of GitHub's Policies.

256.   GitHub's Policies contain multiple explicit provisions that GitHub will not sell the Licensed Materials of the Plaintiffs and Class. GitHub's Terms of Service document provides that the "License Grant to [GitHub] . . . does not grant GitHub the right to sell Your Content." Similarly, GitHub's Privacy Statement defines "personal data" to include "any . . . documents, or other files", a definition that necessarily comprises source code, and hence the Licensed Materials. (As of May 2023, GitHub has updated this provision on its website to explicitly read "any code, text, … documents, or other files"). Elsewhere, the Privacy Statement provides "We do not sell your personal information," "No selling of personal data," "We *do not* sell your personal data for monetary or other consideration." (Emphasis in original).

257.   By making the Licensed Materials available through Copilot in violation of the Suggested Licenses, and charging subscription fees, GitHub has been selling Licensed Materials. By selling the Licensed Materials, GitHub has breached these provisions in GitHub's Policies against selling user data.

258.   GitHub has also breached the implied covenant of good faith and fair dealing. GitHub has long held itself out as a good citizen of the global open-source community. GitHub's Policies were designed to attract Plaintiffs and the Class to become users of the GitHub website by supporting their open-source efforts with fair and ethical practices. By releasing Copilot, GitHub created a product designed to compete with Plaintiffs and the Class and undermine their individual open-source communities. In so doing, GitHub did not act fairly or in good faith.

259.   Plaintiffs and the Class have suffered monetary damages as a result of GitHub's conduct.

260.   GitHub's conduct is causing and, unless enjoined and restrained by this Court, will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully be compensated or measured in money.

261.   As a direct and proximate result of these material breaches by GitHub, Plaintiffs and the Class are entitled to an injunction requiring GitHub to comply with all the terms of the GitHub Policies.

262.   Plaintiffs and the Class are further entitled to recover from GitHub the damages Plaintiffs and the Class sustained—including consequential damages—for Plaintiffs' and the Class's costs in enforcing GitHub's Policies. Plaintiffs and the Class are also entitled to recover as restitution from GitHub for any unjust enrichment, including gains, profits, and advantages that it has obtained as a result of its breaches of the GitHub Policies.

IX. DEMAND FOR JUDGMENT

WHEREFORE, Plaintiffs requests that the Court enter judgment on their behalf and on behalf of the Class defined herein, by adjudging and decreeing that:

263.   This action may proceed as a class action, with Plaintiffs serving as Class Representatives, and with Plaintiffs' counsel as Class Counsel;

   a)   Judgment in favor of Plaintiffs and the Class and against Defendants;

   b)   Permanent injunctive relief, including but not limited to making changes to its Copilot product to ensure that all applicable information set forth in 17 U.S.C. § 1203(b)(1) is included in along with any Output including associated code;

   c)   An order of costs and allowable attorney's fees pursuant to 17 U.S.C. § 1203(b)(4)–(5);

1   d) An award of damages for harms resulting from Defendants' breach of Licenses;

2   e) An award of damages in the amount Defendants have been unjustly enriched through their

3   conduct as alleged herein as well as punitive damages in connection with this conduct;

4   f) An award of damages for harms resulting from GitHub's breach of the GitHub Policies;

5   and

6   264.   Injunctive relief sufficient to alleviate and stop Defendants' unlawful conduct alleged

7 herein.

8   265.   Plaintiffs and the Class are entitled to prejudgment and post-judgment interest on the

9 damages awarded them, and that such interest be awarded at the highest legal rate from and after the date

10 this class action complaint is first served on Defendants;

11   266.   Defendants are to be jointly and severally responsible financially for the costs and expenses

12 of a Court approved notice program through post and media designed to give immediate notification to

13 the Class.

14   267.   Plaintiffs and the Class receive such other or further relief as may be just and proper.

15   **X. JURY TRIAL DEMANDED**

16   Pursuant to Federal Rule of Civil Procedure 38(b), Plaintiffs demand a trial by jury of all the claims

17 asserted in this Complaint so triable.

18

19

20

21

22

23

24

25

26

27

28

1   Dated: January 24, 2024   By: */s/ Joseph R. Saveri*
                                                          Joseph R. Saveri
2
                                                Joseph R. Saveri (State Bar No. 130064)
3   Cadio Zirpoli (State Bar No. 179108)
                                                Christopher K.L. Young (State Bar No. 318371)
4   Louis A. Kessler (State Bar No. 243703)
                                                Elissa A. Buchanan (State Bar No. 249996)
5   Travis Manfredi (State Bar No. 281779)
                                                William W. Castillo Guardado (State Bar No. 294159)
6   Holden J. Benon (State Bar No. 325847)
7   **JOSEPH SAVERI LAW FIRM, LLP**
                                                601 California Street, Suite 1000
8   San Francisco, California 94108
                                                Telephone: (415) 500-6800
9   Facsimile: (415) 395-9940
                                                Email: jsaveri@saverilawfirm.com
10   czirpoli@saverilawfirm.com
11   cyoung@saverilawfirm.com
                                                               lkessler@saverilawfirm.com
12   eabuchanan@saverilawfirm.com
                                                               tmanfredi@saverilawfirm.com
13   wcastillo@saverilawfirm.com
                                                               hbenon@saverilawfirm.com
14
15
                                                Matthew Butterick (State Bar No. 250953)
16   1920 Hillhurst Avenue, #406
                                                Los Angeles, CA 90027
17   Telephone: (323) 968-2632
                                                Facsimile: (415) 395-9940
18   Email: mb@buttoricklaw.com
19
                                                Counsel for Individual and Representative
20   *Plaintiffs and the Proposed Class*
21
22
23
24
25
26
27
28

EXHIBIT E

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

J. DOE 1, et al.,	Case No. 22-cv-06823-JST
Plaintiffs,	
v.	**ORDER GRANTING IN PART AND DENYING IN PART MOTIONS TO DISMISS**
GITHUB, INC., et al.,	
Defendants.	Re: ECF Nos. 50, 53

Before the Court are motions to dismiss filed by Defendants GitHub, Inc. and Microsoft Corporation, ECF No. 50; and Defendants OpenAI, Inc., OpenAI, L.P., OpenAI GP, L.L.C., OpenAI Startup Fund GP I, L.L.C., OpenAI Startup Fund I, L.P., and OpenAI Startup Fund Management, LLC (collectively "OpenAI Defendants"), ECF No. 53. Court will grant the motions in part and deny them in part.

I.   BACKGROUND

Plaintiffs are software developers who challenge Defendants' development and operation of Copilot and Codex, two artificial intelligence-based coding tools.[1] For the purposes of the present motions, the Court accepts as true the following facts in the operative complaint.[2]

GitHub, which was acquired by Microsoft in 2018, is the largest internet hosting service for software projects stored in Git, a widely used open-source version control system for managing

---

[1] Plaintiffs define "artificial intelligence" ("AI") as "a computer program that algorithmically simulates human reasoning or inference, often using statistical methods." Compl. ¶ 2.

[2] By stipulation and order, ECF No. 47, the operative complaint in the consolidated case is that filed in *Doe 3 et al. v. GitHub, Inc. et al.*, No. 22-cv-7074-JST, ECF No. 1. All references to the complaint in this order refer to the operative complaint in the consolidated case.

1   software source code. Using GitHub permits software developers or programmers to collaborate

2   on projects stored in repositories. Repositories may be private or public; anyone can view and

3   access code stored in public repositories.

4   All code uploaded to GitHub is subject to the GitHub Terms of Service, which provide that

5   users retain ownership of any content they upload to GitHub, but grant GitHub the "right to store,

6   archive, parse, and display [the content], and make incidental copies, as necessary to provide the

7   Service, including improving the Service over time." No. 22-cv-7074-JST, ECF No. 1-2 at 27.

8   This "includes the right to do things like copy [the code] to our database and make backups; show

9   it to you and other users; parse it into a search index or otherwise analyze it on our servers; [and]

10   share it with other users." *Id.* at 27-28. Further, the Terms of Service provide that users who set

11   their repositories to be viewed publicly "grant each User of GitHub a nonexclusive, worldwide

12   license to use, display, and perform [the content] through the GitHub Service and to reproduce

13   [the content] solely on GitHub as permitted through GitHub's functionality." *Id.* at 28.

14   Without AI-based assistance, programmers generally write code "both by originating code

15   from the writer's own knowledge of how to write code as well as by finding pre-written portions

16   of code that—under the terms of the applicable license—may be incorporated into the coding

17   project." Compl. ¶ 78. Plaintiffs have each published licensed materials in which they hold a

18   copyright interest to public repositories on GitHub. When creating a new repository, a GitHub

19   user may "select[] one of thirteen licenses from a dropdown menu to apply to the contents of that

20   repository." *Id.* ¶ 34 n.4. Two of the suggested licenses waive copyrights and related rights. The

21   remaining eleven suggested licenses[3] require that any derivative work or copy of the licensed work

22   include attribution to the owner, inclusion of a copyright notice, and inclusion of the license terms.

23   Each Plaintiff published code to a public repository on GitHub under one of the eleven suggested

24

25   [3] These eleven licenses are (1) Apache License 2.0; (2) GNU General Public License version 3
26   ("GPL-3.0"); (3) MIT License; (4) The 2-Clause BSD License ("BSD 2"); (5) The 3-Clause BSD
     License ("BSD 3"); (6) Boost Software License; (7) Eclipse Public License 2.0; (8) GNU Affero
27   General Public License version 3 ("AGPL-3.0"); GNU General Public License version 2 ("GPL
     2"); (10) GNU Lesser General Public License version 2.1 ("LGPL-2.1"); and (11) Mozilla Public
28   License 2.0. Compl. ¶ 34 n.4.

1   licenses that include these three requirements.

2   In June 2021, GitHub and OpenAI released Copilot, an AI-based program that can "assist

3   software coders by providing or filling in blocks of code using AI." *Id.* ¶ 8. In August 2021,

4   OpenAI released Codex, an AI-based program "which converts natural language into code and is

5   integrated into Copilot." *Id.* ¶ 9. Codex is integrated into Copilot: "GitHub Copilot uses the

6   OpenAI Codex to suggest code and entire functions in real-time, right from your editor." *Id.* ¶ 47

7   (quoting GitHub website). GitHub users pay $10 per month or $100 per year for access to

8   Copilot. *Id.* ¶ 8.

9   Codex and Copilot employ machine learning, "a subset of AI in which the behavior of the

10   program is derived from studying a corpus of material called training data." *Id.* ¶ 2. Using this

11   data, "through a complex probabilistic process, [these programs] predict what the most likely

12   solution to a given prompt a user would input is." *Id.* ¶ 79. Codex and Copilot were trained on

13   "billions of lines" of publicly available code, including code from public GitHub repositories. *Id.*

14   ¶¶ 82-83.

15   Despite the fact that much of the code in public GitHub repositories is subject to open-

16   source licenses which restrict its use, *id.* ¶ 20, Codex and Copilot "were not programmed to treat

17   attribution, copyright notices, and license terms as legally essential," *id.* ¶ 80. Copilot reproduces

18   licensed code used in training data as output with missing or incorrect attribution, copyright

19   notices, and license terms. *Id.* ¶¶ 56, 71, 74, 87-89. This violates the open-source licenses of

20   "tens of thousands—possibly millions—of software developers." *Id.* ¶ 140. Plaintiffs additionally

21   allege that Defendants improperly used Plaintiffs' "sensitive personal data" by incorporating the

22   data into Copilot and therefore selling and exposing it to third parties. *Id.* ¶¶ 225-39.

23   Plaintiffs filed multiple cases against Defendants, which were subsequently consolidated.

24   ECF No. 47. Plaintiffs, on behalf of themselves and two putative classes,[4] plead twelve counts

25

26   [4] The "Injunctive Relief Class" and "Damages Class" are each defined as: "All persons or entities
     domiciled in the United States that[] (1) owned an interest in at least one [U.S.] copyright in any
27   work; (2) offered that work under one of GitHub's Suggested Licenses; and (3) stored Licensed
     Materials in any public GitHub repositories at any time between January 1, 2015 and the present
28   (the 'Class Period')." Compl. ¶ 34 (footnote omitted).

United States District Court
Northern District of California

1   against Defendants: (1) violation of the Digital Millennium Copyright Act ("DMCA"), 17 U.S.C.

2   §§ 1201-05; (2) common law breach of open-source licenses; (3) common law tortious

3   interference in a contractual relationship; (4) common law fraud; (5) false designation of origin in

4   violation of the Lanham Act, 15 U.S.C. § 1125; (6) unjust enrichment in violation of Cal. Bus. &

5   Prof. Code §§ 17200, *et seq.*, and the common law; (7) unfair competition in violation of the

6   Lanham Act, 15 U.S.C. § 1125; Cal. Bus. & Prof. Code §§ 17200, *et seq.*, and the common law;

7   (8) breach of contract for violation of the GitHub Privacy Policy and Terms of Service; (9)

8   violation of the California Consumer Privacy Act ("CCPA"); (10) common law negligence; (11)

9   common law civil conspiracy; and (12) declaratory relief under 28 U.S.C. § 2201(a) and Cal. Code

10   Civ. Proc. § 1060.[5] Defendants now move to dismiss the complaint. ECF Nos. 50, 53.

11   **II.   JURISDICTION**

12   The Court has jurisdiction over Plaintiffs' federal claims under 28 U.S.C. § 1331 and

13   supplemental jurisdiction over Plaintiffs' state law claims under 28 U.S.C. § 1367.

14   **III.   LEGAL STANDARD**

15   **A.   Rule 12(b)(1)**

16   "Article III of the Constitution confines the federal judicial power to the resolution of

17   'Cases' and 'Controversies.'" *TransUnion LLC v. Ramirez*, 141 S. Ct. 2190, 2203 (2021). "For

18   there to be a case or controversy under Article III, the plaintiff must have a 'personal stake' in the

19   case—in other words, standing." *Id.* (quoting *Raines v. Byrd*, 521 U.S. 811, 819 (1997)). A

20   defendant may attack a plaintiff's assertion of jurisdiction by moving to dismiss under Rule

21   12(b)(1) of the Federal Rules of Civil Procedure. *Cetacean Cmty. v. Bush*, 386 F.3d 1169, 1174

22

23   [5] While Plaintiffs plead several common law tort claims, they do not identify the state law which
     applies to each claim. "[D]ue to variances among state laws, failure to allege which state law
24   governs a common law claim is grounds for dismissal." *In re Nexus 6P Prods. Liab. Litig.*, 293 F.
     Supp. 3d 888, 933 (N.D. Cal. 2018) (quoting *Romero v. Flowers Bakeries, LLC*, No. 14-cv-05189-
25   BLF, 2016 WL 469370, at *12 (N.D. Cal. Feb. 8, 2016)). With respect to the state law claims in
     any future amended complaint, Plaintiffs shall identify the state under whose law the claim is
26   brought. When claims which share a legal theory are brought under multiple laws (for example, a
     federal statute, a state statute, and common law, such as Plaintiffs' claim for unfair competition),
27   Plaintiffs shall state each claim as a separate count. For the purposes of deciding the present
     motion to dismiss, the Court interprets the complaint as asserting the common law claims of
28   Counts 2, 3, 4, 6, 7, 10, and 11 under California law.

1   (9th Cir. 2004); *see also Maya v. Centex Corp.*, 658 F.3d 1060, 1067 (9th Cir. 2011) ("[L]ack of

2   Article III standing requires dismissal for lack of subject matter jurisdiction under Federal Rule of

3   Civil Procedure 12(b)(1).").

4   "A Rule 12(b)(1) jurisdictional attack may be facial or factual." *Safe Air for Everyone v.*

5   *Meyer*, 373 F.3d 1035, 1039 (9th Cir. 2004). "In a facial attack, the challenger asserts that the

6   allegations contained in a complaint are insufficient on their face to invoke federal jurisdiction."

7   *Id.* Where, as here, a defendant makes a facial attack, the court assumes that the complaint's

8   allegations are true and draws all reasonable inferences in the plaintiff's favor. *Wolfe v.*

9   *Strankman*, 392 F.3d 358, 362 (9th Cir. 2004).

10   **B.   Rule 12(b)(6)**

11   "Dismissal under [Federal Rule of Civil Procedure] 12(b)(6) is appropriate only where the

12 complaint lacks a cognizable legal theory or sufficient facts to support a cognizable legal theory."

13 *Mendiondo v. Centinela Hosp. Med. Ctr.*, 521 F.3d 1097, 1104 (9th Cir. 2008). A complaint must

14 contain "a short and plain statement of the claim showing that the pleader is entitled to relief."

15 Fed. R. Civ. P. 8(a)(2). Facts pleaded by a plaintiff "must be enough to raise a right to relief

16 above the speculative level." *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 555 (2007).

17   "To survive a motion to dismiss, a complaint must contain sufficient factual matter,

18 accepted as true, to 'state a claim to relief that is plausible on its face.'" *Ashcroft v. Iqbal*, 556 U.S.

19 662, 678 (2009) (quoting *Twombly*, 550 U.S. at 570)). "A claim has facial plausibility when the

20 plaintiff pleads factual content that allows the court to draw the reasonable inference that the

21 defendant is liable for the misconduct alleged." *Id.* In determining whether a plaintiff has met this

22 plausibility standard, the Court must "accept all factual allegations in the complaint as true and

23 construe the pleadings in the light most favorable to the nonmoving party." *Knievel v. ESPN*, 393

24 F.3d 1068, 1072 (9th Cir. 2005).

25   **C.   Leave to Amend**

26   Leave to amend a complaint "shall be freely given when justice so requires." Fed. R. Civ.

27 P. 15(a)(2). The decision of whether to grant leave to amend is "within the discretion of the

28 district court, which may deny leave due to 'undue delay, bad faith or dilatory motive on the part

United States District Court
Northern District of California

5

1   of the movant, repeated failure to cure deficiencies by amendments previously allowed, undue

2   prejudice to the opposing party by virtue of allowance of the amendment, and futility of

3   amendment.'" *Leadsinger, Inc. v. BMG Music Pub.*, 512 F.3d 522, 532 (9th Cir. 2008) (quoting

4   *Foman v. Davis*, 371 U.S. 178, 182 (1962)).

5   **IV.   DISCUSSION**

6   **A.   Article III Standing**

7   Defendants argue that Plaintiffs lack Article III standing to assert their claims and that the

8   Court therefore lacks subject-matter jurisdiction over this action.[6]

9   "[T]o establish standing, a plaintiff must show (i) that he suffered an injury in fact that is

10   concrete, particularized, and actual or imminent; (ii) that the injury was likely caused by the

11   defendant, and (iii) that the injury would likely be redressed by judicial relief." *TransUnion*, 141

12   S. Ct. at 2203. "The party invoking federal jurisdiction bears the burden of establishing these

13   elements." *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 561 (1992). "Where, as here, a case is at

14   the pleading stage, the plaintiff must 'clearly . . . allege facts demonstrating' each element."

15   *Spokeo, Inc. v. Robins*, 578 U.S. 330, 338 (2016) (quoting *Warth v. Seldin*, 422 U.S. 490, 518

16   (1975)).

17   "A plaintiff must demonstrate constitutional standing separately for each form of relief

18   requested." *Davidson v. Kimberly-Clark*, 889 F.3 956, 969 (9th Cir. 2018). In a putative class

19   action, named plaintiffs "must allege and show that they personally have been injured, not that

20   injury has been suffered by other, unidentified members of the class to which they belong and

21   which they purport to represent." *Warth*, 422 U.S. at 502.

22   Plaintiffs advance two main theories of harm. One theory is that Defendants have shared,

23

---

24   [6] Both parties impermissibly incorporate substantive material by reference across their briefs.
"The incorporation of substantive material by reference is not sanctioned by the federal rules."
25   *Swanson v. U.S. Forest Serv.*, 87 F.3d 339, 345 (9th Cir. 1996). In the future, all arguments the
parties seek to make must be raised in the briefing on a particular motion. *See Woolfson v. Conn*
26   *Appliances, Inc.*, No. 21-cv-07833-MMC, 2022 WL 3139522, at *6 (N.D. Cal. Aug. 5, 2022)
(declining to consider substantive arguments purportedly incorporated by reference); *Williams v.*
27   *Cnty. of Alameda*, 26 F. Supp. 3d 925, 947 (N.D. Cal. 2014) ("[T]he Court will not consider the
arguments that Plaintiff improperly seeks to incorporate by reference. This Court only considers
28   arguments that are specifically and distinctively raised by the parties in their briefs.").

United States District Court
Northern District of California

1   sold, and exposed and will continue to share, sell, and expose Plaintiffs' personal information,

2   harming Plaintiffs' rights to privacy and the protection of their personal data. The second is that

3   Defendants' use of licensed code as training data for Codex and Copilot harms Plaintiffs' property

4   interests in the licensed code because Copilot already has or likely will reproduce Plaintiffs' code

5   in violation of those licenses.

6   **1.   Injury to Privacy Rights**

7   Defendants argue that Plaintiffs fail to plausibly allege any disclosure of personal

8   information, and therefore fail to allege an actual or imminent injury sufficient to confer standing.

9   In the complaint, Plaintiffs do not identify the specific sensitive or private information at issue.

10   Plaintiffs therefore do not allege facts sufficient for the Court to infer that Plaintiffs' privacy

11   interests are implicated by the alleged misuse of such information. *See Ji v. Naver Corp.*, No. 21-

12   cv-05143-HSG, 2022 WL 4624898, at *7 (N.D. Cal. Sept. 30, 2022) (dismissing privacy-based

13   claim for lack of standing where plaintiffs "have not alleged enough facts to show that the

14   [personal data at issue] are the type of information that could give rise to a privacy injury"); *I.C. v.*

15   *Zynga, Inc.*, 600 F. Supp. 3d 1034, 1049 (N.D. Cal. Apr. 29, 2022) (evaluating whether privacy

16   interests were implicated by allegedly disclosed personal information to determine whether

17   plaintiffs had standing).

18   Plaintiffs have not met their burden to allege facts demonstrating an injury-in-fact

19   sufficient to confer standing for their privacy-based claims. Plaintiffs' claims for breach of the

20   GitHub Privacy Policy and Terms of Service, violation of the CCPA, and negligence are

21   dismissed with leave to amend.

22   **2.   Injury to Property Rights**

23   Defendants argue that Plaintiffs do not identify any instance of Copilot reproducing

24   Plaintiffs' licensed code and therefore fail to plead a particularized injury sufficient to confer

25   standing.

26   "For an injury to be 'particularized,' it 'must affect the plaintiff in a personal and

27   individual way.'" *Spokeo*, 578 U.S. at 339 (quoting *Lujan*, 504 U.S. at 560 n.1). "[A]t an

28   irreducible minimum, Art. III requires the party who invokes the court's authority to 'show that he

1   personally has suffered some actual or threatened injury.'"   *Valley Forge Christian Coll. v. Ams.*

2   *United for Separation of Church and State, Inc.*, 454 U.S. 464, 472 (1982) (quoting *Gladstone*

3   *Realtors v. Village of Bellwood*, 441 U.S. 91, 99 (1979)).

4   Plaintiffs argue that they "have alleged the requisite particularized injury" because they

5   "allege that Defendants have violated provisions of [the] open-source licenses" under which

6   Plaintiffs published their code. ECF No. 67 at 14. In the complaint, Plaintiffs allege that

7   Defendants violate those licenses because Copilot reproduces Plaintiffs' code as output with

8   missing or incorrect attribution, copyright notices, and license terms.[7] However, while Plaintiffs

9   identify several instances in which Copilot's output matched licensed code written by a Github

10   user, Compl. ¶¶ 56, 71, 74, 87-89, none of these instances involve licensed code published to

11   GitHub by Plaintiffs. Because Plaintiffs do not allege that they themselves have suffered the

12   injury they describe, they do not have standing to seek retrospective relief for that injury.[8]

13   Plaintiffs also argue that they "also allege an imminent and significant harm—now that

14   their software code is part of the training data . . . , [it] can be output at any time without the

15   legally necessary notices and attribution, leading to increased and imminent risk of

16   misappropriation." ECF No. 67 at 15. To the extent that Plaintiffs allege an injury based on the

17   increased risk of misappropriation of their own licensed code, that injury is sufficiently

18   particularized. However, an increased risk of future harm alone is not sufficiently concrete to

19   confer standing for damages. *TransUnion*, 141 S. Ct. at 2210-11 (finding "persuasive" argument

20

21   [7] In opposition, Plaintiffs suggest they were also injured by Defendants' use of Plaintiffs' licensed
     code as training data for Codex and Copilot. ECF No. 66 at 12, 21. But Plaintiffs' complaint does
22   not describe such an injury. In the complaint, Plaintiffs allege that "Defendants had access to but
     were not licensed by Plaintiffs . . . to train any . . . functional prediction engine using the Licensed
23   Materials." Compl. ¶ 149; *see also id.* ¶ 37 (describing one question common to the putative class
     as "[w]hether Defendants violated the Licenses governing use of the Licensed Materials by using
24   them to train Copilot"). However, Plaintiffs do not actually allege that Defendants' use of
     licensed code to train Codex and Copilot constituted a breach of the open-source licenses at issue.
25   *See id.* ¶¶ 172-87 (alleging that Defendants breached the licenses by failing to provide attribution
     in output, failing to provide copyright notices in output, and failing to identify applicable licenses
26   and the text of those licenses in output). Because Plaintiffs do not allege they were injured by
     Defendants' use of licensed code as training data, the Court does not address whether such an
27   injury is sufficient to confer standing.

28   [8] This problem is related to redressability: Plaintiffs allege no injury redressable by the monetary
     damages they seek.

8

1   that, "in a suit for damages, the mere risk of future harm . . . cannot qualify as a concrete harm . . .

2   unless the exposure to the risk of future harm itself causes a *separate* concrete harm") (emphasis

3   in original); *id.* at 2213 (holding that "the risk of future harm on its own does not support Article

4   III standing for the plaintiffs' damages claims"). Plaintiffs do not allege any additional, concrete

5   harm associated with this increased risk of misappropriation. Thus, an increased risk of

6   misappropriation cannot provide standing for Plaintiffs' damages claims.

7   Finally, Plaintiffs argue that, "[g]iven the number of times users may use Copilot, it is a

8   virtual certainty [that] any particular plaintiff's code will be displayed either with copyright

9   notices removed or in violation of Plaintiffs' open-source licenses for profit." ECF No. 67 at 15.

10   Though this does not support standing for retrospective damages – for which, as discussed above,

11   Plaintiffs must allege that they have suffered a qualifying injury-in-fact that has actually occurred

12   to them – it may support standing for injunctive relief.

13   "[A] person exposed to a risk of future harm may pursue forward-looking, injunctive relief

14   to prevent the harm from occurring, at least so long as the risk of harm is sufficiently imminent

15   and substantial." *TransUnion*, 141 S. Ct. at 2210. "An allegation of future injury may suffice if

16   the threatened injury is 'certainly impending,' or there is a 'substantial risk that the harm will

17   occur.'" *Susan B. Anthony List v. Driehaus*, 573 U.S. 149, 158 (2014) (quoting *Clapper v.

18   Amnesty Int'l USA*, 568 U.S. 414 n.5 (2013)). A substantial risk means there is a "realistic danger

19   of sustaining a direct injury." *Pennell v. City of San Jose*, 485 U.S. 1, 8 (1988) (quoting *Babbitt v.

20   United Farm Workers Nat'l Union*, 442 U.S. 289, 298 (1979)).

21   Plaintiffs plausibly allege that there is at least a substantial risk that Defendants' programs

22   will reproduce Plaintiffs' licensed code as output. Plaintiffs allege that both Codex and Copilot

23   were trained on data that included all public GitHub repositories. Plaintiffs further allege that the

24   programs reproduce well-known code in response to related prompts and offer several concrete

25   examples of such reproduction. Finally, Plaintiffs allege that GitHub's own internal research

26   shows that Copilot reproduces code from training data "about 1% of the time." Compl. ¶¶ 56, 71,

27   74, 87-90. Plaintiffs thus plausibly allege that, absent injunctive relief, there is a realistic danger

28   that Codex or Copilot will reproduce Plaintiffs' licensed code as output. Plaintiffs further allege

1   that Defendants have modified Copilot to ensure that it no longer reproduces license text,

2   attribution, and copyright notices. Taking the facts of the complaint as true and construing all

3   inferences in the Plaintiffs' favor, the Court can reasonably infer that, should Plaintiffs' code be

4   reproduced as output, it will be reproduced in a manner that violates the open-source licenses

5   under which Plaintiffs published their code.[9]

6   While Plaintiffs have failed to establish an injury-in-fact sufficient to confer standing for

7   their claims for damages based on injury to property rights, they have standing to pursue

8   injunctive relief on such claims.

9   **B.   Proceeding Pseudonymously**

10   Defendants argue that the complaint should be dismissed because Plaintiffs are not entitled

11   to proceed under fictitious names.[10]

12   "The normal presumption in litigation is that parties must use their real names." *Doe v.*

13   *Kamehameha Schs./Bernice Pauahi Bishop Est.*, 596 F.3d 1036, 1042 (9th Cir. 2010). Parties

14   may proceed pseudonymously only "in the 'unusual case' when nondisclosure of the party's

15   identity 'is necessary . . . to protect a person from harassment, injury, ridicule or personal

16   embarrassment." *Does I Thru XXIII v. Advanced Textile Corp.*, 214 F.3d 1058, 1067-68 (9th Cir.

17   2000) (alteration in original) (quoting *United States v. Doe*, 655 F.2d 920, 922 n.1 (9th Cir.

18   1981)).

19   Threats of physical harm "present[] the paradigmatic case for allowing anonymity."

20   *Kamehameha*, 596 F.3d at 1043. Where a party seeks to proceed pseudonymously on the basis of

21

22   [9] Defendants argue that Plaintiffs must also allege facts regarding the type of code they published,
     what problem that code solves, how frequently that code might appear on GitHub, and how likely
23   it is that a user would enter a prompt that would generate a match to that code. ECF No. 50 at 16.
     Defendants suggest that, absent such facts, Plaintiffs' alleged injury is too "conjectural or
24   hypothetical" to confer standing. *Id.* (quoting *Lujan*, 504 U.S. at 550). Defendants ask too much
     of Plaintiffs at the pleading stage. Such facts might help Defendants—the only parties with
25   knowledge of how Copilot and Codex were designed and operate—understand whether Plaintiffs'
     alleged injury has already occurred or how soon it is likely to occur. However, taking the facts in
26   the complaint as true, and construing all inferences in Plaintiffs' favor, their alleged future injury
     is neither conjectural nor hypothetical: their licensed code was used to train these programs and
27   there is presently a realistic danger that the code will be reproduced as output.

28   [10] Because the caselaw uses "pseudonymous" and "anonymous" interchangeably, this Court does
     so as well.

United States District Court
Northern District of California

1
2
3
4

retaliatory harm, "a district court must balance five factors: '(1) the severity of the threatened

harm, (2) the reasonableness of the anonymous party's fears, . . . (3) the anonymous party's

vulnerability to such retaliation,' (4) the prejudice to the opposing party, and (5) the public

interest." *Id.* at 1042 (quoting *Advanced Textile*, 214 F.3d at 1068).

5
6
7
8
9
10
11
12
13
14
15
16
17

Plaintiffs seek to proceed pseudonymously because, "through their counsel, [they] have

received legitimate and credible threats of physical violence." ECF No. 67 at 17. Plaintiffs

highlight three threatening emails sent to counsel prior to the filing of the operative complaint. All

three emails contain veiled threats wishing death upon Plaintiffs' counsel on the basis of his

involvement in this lawsuit. ECF No. 68-1 at 2 ("[I]magine shooting against [AI] . . . . [G]o kys. .

. . [K]ill urself."); ECF No. 68-2 at 2 ("I hope you f*cking die you piece of sh*t. It's people like

you why this world sucks so f*cking bad and we can't have nice things. I literally hope someone

murder [sic] you. Go f*cking die."); ECF No. 68-3 at 2 ("[G]o f*cking cry about [G]it[H]ub you

f*cking piece of sh*t n****r, [I] hope your throat gets cut open and every single family member

of you [sic] is burnt to death."). Plaintiffs' counsel have received "many" such messages. ECF

No. 67 at 17. Plaintiffs argue they should be permitted to proceed pseudonymously because they

reasonably fear they will be subject to threats of retaliation and violence if their identities are

disclosed.

18
19
20
21
22
23
24

The first two factors of the balancing test for retaliatory harm – severity of the threatened

harm and the reasonableness of the fear of such harm – "are intricately related and should be

addressed together." *Kamehameha*, 596 F.3d at 1040. Plaintiffs need not prove the speakers

"intend to carry out the threatened retaliation," but rather "that a reasonable person would believe

that the threat might actually be carried out." *Advanced Textile*, 214 F.3d at 1071. "It is in the

particular purview of the district court to view alleged threats in context and determine what the

'reasonable' person in the plaintiffs' situation would fear." *Kamehameha*, 596 F.3d at 1044. "In

25
26
27
28

11

1   context, a plaintiff might reasonably fear a veiled threat of violence." *Id.*[11]

2   The threatened harm in this case – death – is plainly severe. *Id.* at 1043 (describing threats

3   of physical retaliation, including death, as "undoubtedly severe"); *Doe v. Steagall*, 653 F.2d 180,

4   186 (5th Cir. 1981) (permitting plaintiffs to proceed pseudonymously where they faced "threats of

5   violence," including veiled death threats, "generated by this case"); *Doe v. Univ. Acct. Serv., LLC*,

6   No. 09-CV-01563-BAS-JLB, 2022 WL 623913, at *4 (S.D. Cal. Mar. 3, 2022) (permitting

7   plaintiff to proceed pseudonymously where he received death threats). While the threatening

8   emails were not sent to Plaintiffs directly, the emails wish death upon Plaintiffs' counsel on the

9   basis of their involvement in this lawsuit. It is reasonable for Plaintiffs to fear that such threats

10 might be carried out against them if their identities were to become public.

11   GitHub and Microsoft suggest that Plaintiffs' fear is unreasonable because "the types of

12 nasty messages at issue here are a fact of modern life in the era of internet 'trolls.'" ECF No. 72 at

13 19. GitHub and Microsoft do not explain why the rise of internet trolls renders Plaintiffs' fears of

14 harm unreasonable. *Cf. United States v. Bagdasarian*, 652 F.3d 1113, 1126-27 (9th Cir. 2011)

15 (Wardlaw, J., concurring in part) (describing the connection between anonymous internet posts

16 and subsequent real-world violence). Sending direct messages containing veiled death threats

17 would seem to constitute behavior beyond trolling. *See* Merriam Webster's Collegiate Dictionary

18 1341 (11th ed. 2003) (defining "troll," in relevant part, as "to antagonize (others) online by

19 *posting* inflammatory, irrelevant, or offensive comments or content" and "to harass, criticize, or

20 antagonize (someone) esp[ecially] by provocatively disparaging or mocking *public* statements,

21 postings, or acts") (emphasis added). These were not public posts mocking or antagonizing

22 Plaintiffs' counsel, but rather private emails, sent directly to Plaintiffs' counsel, wishing him and

23 his family violent death. Where many individuals take the time to send private, threatening

24

25 [11] Of note, the Ninth Circuit has explained that the standard for reasonable fear in this context is

26 more permissive than in First Amendment "true threat" cases. *Kamehameha*, 596 F.3d at 1044 n.6
    ("The First Amendment cases discussing the concept of a 'true threat' . . . pose a higher bar to

27 finding a reasonable fear. In those cases, one party's fear of the threat must be weighed against
    the opposing party's first amendment right to speak freely because the threatened party seeks to

28 prevent the other party's speech.").

1   emails, it is reasonable to fear that some of those individuals might carry out their threats. The

2   Court finds that the first two factors weigh in favor of permitting Plaintiffs to proceed

3   pseudonymously.

4   Defendants identify no prejudice from Plaintiffs proceeding pseudonymously. The Ninth

5   Circuit has explained that, where a defendant does not know the plaintiff's name, "at some later

6   point in the proceedings it may be necessary to reveal plaintiffs' identities to defendants so that

7   defendants may refute [their] individualized accusations." *Advanced Textile*, 214 F.3d at 1058.

8   "But where the defendants know the plaintiffs' names, 'anonymity need not, and should not,

9   impede either party's ability to develop its case' even though it is 'foreseeable that anonymity

10 [will] raise problems for discovery.'" *Doe 1 v. Nat'l Collegiate Athletic Assoc.*, No. 22-cv-01559-

11 LB, 2022 WL 3974098, at *2 (N.D. Cal. Aug. 30, 2022) (quoting *Jane Roes 1-2 v. SFBSC Mgmt.,*

12 *LLC*, 77 F. Supp. 3d 990, 996 (N.D. Cal. 2015)). Plaintiffs have disclosed their true names to

13 Defendants subject to a protective order, so pseudonymity should not impede Defendants' ability

14 to develop their case. Though pseudonymity may pose certain logistical challenges during

15 discovery, this case remains at the pleadings stage. *See Doe v. County of El Dorado*, No. 2:13-

16 CV-01433-KJM, 2013 WL 6230342, at *5 (E.D. Cal. Dec. 2, 2013) (explaining that, "at the

17 prediscovery stage . . . the court need not yet consider the prejudice defendant will suffer during

18 discovery," as "the relevant prejudice is that which defendant presently suffers as a result of

19 plaintiff's anonymity"). Defendants do not articulate any prejudice associated with Plaintiffs

20 proceeding pseudonymously at this stage of the litigation, and this Court is not aware of any.

21 Therefore, the Court finds that the prejudice factor does not weigh against permitting Plaintiffs to

22 proceed pseudonymously at this stage.

23   Finally, the public interest factor does not weigh against permitting Plaintiffs to proceed

24 pseudonymously. "The normal presumption in litigation . . . that parties must use their real names

25 . . . is loosely related to the public's right to open courts and the right of private individuals to

26 confront their accusers." *Kamehameha*, 596 F.3d at 1042. Where the plaintiffs' identities are not

27 central to the issues raised by a case, however, the public interest may not be harmed by

28 permitting plaintiffs to proceed pseudonymously. *See Advanced Textile* , 214 F.3d at 1072

1 (reversing denial of anonymity where "[t]he district court did not explain, and we fail to see, how

2 disguising plaintiffs' identities will obstruct public scrutiny of the important issues in th[e] case");

3 *Kamehameha*, 596 F.3d at 1043 (noting that, where plaintiffs brought "claims of widespread

4 discrimination," "it [wa]s difficult to see 'how disguising plaintiffs' identities w[ould] obstruct

5 public scrutiny of the important issues in th[e] case'") (quoting *Advanced Textile*, 214 F.3d at

6 1072). Withholding the true identities of the individual software developers who bring this case

7 will not obstruct public scrutiny of the issues raised. Plaintiffs' names and identities have no

8 bearing on the central issues of this case, including whether Defendants' AI-based coding tools

9 illegally reproduce licensed code used as training data. The Court finds that the public interest

10 factor does not weigh against anonymity at this stage of the litigation. *See El Dorado*, 2013 WL

11 6230342, at *6 (finding public interest factor weighed in favor of anonymity where the

12 "[p]laintiff's identity appears to have no bearing on the resolution of the issues, and a pseudonym

13 will not impede public access to the substance of the proceedings").[12]

14   Balancing these factors, the Court finds that Plaintiffs may proceed pseudonymously at this

15 time.

16   **C.**   **Rule 8(a)**

17   Defendants argue that the complaint should be dismissed under Federal Rule of Civil

18 Procedure 8(a) because Plaintiffs do not plead sufficient facts regarding the role of each Defendant

19 in the alleged misconduct.

20   A complaint must contain "a short and plain statement of the claim showing that the

21 pleader is entitled to relief." Fed. R. Civ. P. 8(a)(2). "[T]he 'short and plain statement' must

22 provide the defendant with 'fair notice of what the plaintiff's claim is and the grounds upon which

23 it rests.'" *Dura Pharms., Inc. v. Broudo*, 544 U.S. 336, 346 (2005) (quoting *Conley v. Gibson*, 355

24 U.S. 41, 47 (1957)). To comply with Rule 8(a), a plaintiff "must allege the basis of his claim

25

26 ───────────────

[12] The Court is mindful that, like the other factors, the public interest in Plaintiffs' identities may
27 change as the suit progresses. *See Doe v. NFL Enters., LLC*, No. C 17-00496 WHA, 2017 WL
697420, at *2 (N.D. Cal. Feb. 22, 2017) (noting that "class members will . . . have a right to know
28 the identity of their representative in this litigation" in later stages of the case).

1   against each defendant . . . to put defendants on sufficient notice of the allegations against them."

2   *Gauvin v. Trombatore*, 682 F. Supp. 1067, 1071 (N.D. Cal. 1988).

3   The OpenAI Defendants argue that Plaintiffs plead claims against all six OpenAI entities

4   without specifying the role or conduct of each entity, thus violating Rule 8(a). But Plaintiffs

5   allege that two of the OpenAI Defendants – OpenAI, Inc. and OpenAI, L.P. – "programmed,

6   trained, and maintain[]" Codex, and that OpenAI, Inc. owns and controls all of the other OpenAI

7   Defendants. Compl. ¶ 23. Plaintiffs plead additional specific facts regarding the ownership and

8   control of each of the OpenAI Defendants. *See, e.g.*, *id.* ¶ 24 ("OpenAI, L.P. is a wholly owned

9   subsidiary of OpenAI, Inc. . . . [and] the OpenAI entity that co-created Copilot and offers it jointly

10   with GitHub."); *id.* ¶ 25 ("OpenAI GP is the general partner of OpenAI, L.P. OpenAI GP

11   manages and operates the day-to-day business and affairs of OpenAI, L.P. . . . OpenAI GP was

12   aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P."). Such

13   allegations sufficiently put each OpenAI entity on notice of the basis on which Plaintiffs allege it

14   may be liable for the challenged conduct.[13]

15   Similarly, Microsoft and GitHub argue that Plaintiffs' complaint fails to satisfy Rule 8(a)

16   because Plaintiffs do not allege which specific acts taken by each Defendant violate the DMCA.

17   But Plaintiffs allege facts which sufficiently provide notice of the basis on which each GitHub and

18   Microsoft are sued. Microsoft is alleged to maintain an ownership interest in GitHub and OpenAI,

19   L.P., which allegedly co-created Copilot. Compl. ¶ 22 ("Microsoft owns and operates GitHub.

20   Through its corporate ownership, control of the GitHub Board of Directors, active management,

21

22   _____

23   [13] The Court does not decide here whether such alleged relationships in fact provide a sufficient
     basis for liability against each of the OpenAI entities for any of the pleaded claims, because that
     issue is not presently before the Court. The function of Rule 8 in this context is to "give

24   Defendants fair notice of the allegations against them," *Ketayi v. Health Enrollment Grp.*, 516 F.
     Supp. 3d 1092, 1120 (S.D. Cal. 2021) (citations omitted), whereas the purpose of a Rule 12(b)(6)

25   motion is to "test[] the legal sufficiency of the pleadings and allow[] a court to dismiss a complaint

26   upon a finding that the plaintiff has failed to state a claim upon which relief may be granted,"
     Lessin v. Ford Motor Co., 600 F. Supp. 3d 1137, 1141 (S.D. Cal. 2022) (citation omitted). *See*

27   *also Olson v. Puckett*, No. 221CV01482KJMDMC, 2023 WL 2602174, at *1 (E.D. Cal. Mar. 22,
     2023) ("To the extent there is a difference between the standards imposed by Rule 8 and Rule

28   12(b)(6), this court interprets the Rule 8 requirement as the lesser of the two.").

United States District Court
Northern District of California

1   and other means, Microsoft sells, markets and distributes Copilot."); *id.* ¶ 7 ("Microsoft obtained a

2   partial ownership interest in OpenAI . . . . As OpenAI's largest investor . . . Microsoft exerts

3   considerable control over OpenAI."); *id.* ¶ 24 ("OpenAI, L.P. is the OpenAI entity that co-created

4   Copilot and offers it jointly with GitHub."). Plaintiffs further allege, on information and belief,

5   that "Microsoft, GitHub, [and] OpenAI . . . have worked together to create Copilot." *Id.* ¶ 241.

6   Such allegations provide sufficient notice of the basis of Plaintiffs' DMCA claims against each

7   Defendant.[14] The complaint is not deficient under Rule 8(a).

8   **D.   Rule 12(b)(6)**

9   Defendants move to dismiss most of Plaintiffs' claims for failure to state a claim on which

10   relief may be granted. In opposition, Plaintiffs do not address Defendants' substantive arguments

11   about the sufficiency of Plaintiffs' claims for violation of Section 1202(a) of the DMCA, tortious

12   interference, fraud, false designation of origin, or violation of the CCPA. Plaintiffs instead state

13   that they "do not concede" these claims "have been inadequately pled," but request leave to amend

14   the claims. ECF No. 66 at 30; ECF No. 67 at 32 n.17. Without briefing from both parties, the

15   Court will not evaluate the merits of Defendants' arguments as to those claims.

16   Mindful that leave to amend should be freely granted, the Court dismisses Plaintiffs'

17   claims for violation of Section 1202(a) of the DMCA, tortious interference, fraud, false

18   designation of origin, and violation of the CCPA with leave to amend.

19   **1.   Copyright Preemption**

20   Defendants argue that several of Plaintiffs' state law claims are preempted by Section 301

21   of the Copyright Act. Because most of these claims were previously dismissed, the Court only

22   considers whether Plaintiffs' unjust enrichment claim is subject to copyright preemption.

23   The Copyright Act of 1976 expressly preempts state claims where the plaintiff's work

24   "come[s] within the subject matter of copyright" and the state law grants "legal or equitable rights

25   that are equivalent to any of the exclusive rights within the general scope of copyright." 17 U.S.C.

26   § 301(a). The Ninth Circuit has established a two-part test to determine whether state law claims

27

28   [14] Again, the Court does not determine whether such allegations are sufficient to plead a basis for
liability against any Defendant. *See supra* note 13.

United States District Court
Northern District of California

1   are preempted. First, "[the court must] decide 'whether the "subject matter" of the state law claim

2   falls within the subject matter of copyright as described in 17 U.S.C. §§ 102 and 103.'" *Maloney*

3   *v. T3Media, Inc.*, 853 F.3d 1004, 1010 (9th Cir. 2017). If it does, the court must "determine

4   'whether the rights asserted under state law are equivalent to the rights contained in 17 U.S.C.

5   § 106, which articulates the exclusive rights of copyright holders.'" *Id.* "If a state law claim

6   includes an 'extra element' that makes the right asserted qualitatively different from those

7   protected under the Copyright Act, the state law is not preempted by the Copyright Act." *Altera*

8   *Corp. v. Clear Logic, Inc.*, 424 F.3d 1079, 1089 (9th Cir. 2005).

9   Plaintiffs' opposition generally does not address Defendants' preemption arguments.[15]

10   However, Plaintiffs do argue generally that their state law claims are qualitatively different from

11   claims under the Copyright Act because they are not solely about the unauthorized reproduction of

12   Plaintiffs' code, but also the unauthorized use of such code. Plaintiffs suggest that state law tort

13   claims concerning unauthorized use are not preempted by the Copyright Act, and that "Plaintiffs

14   allege that Defendants, through their unauthorized *use* of Plaintiffs' code to train Codex and

15   Copilot, and their display of Plaintiffs' code to others for commercial gain, violated Plaintiffs'

16   rights under state and common law." ECF No. 66 at 18 (emphasis added).

17   Plaintiffs are correct that state law tort claims concerning unauthorized use are not

18   preempted by the Copyright Act. "The [exclusive] rights protected under the Copyright Act

19   include the rights of reproduction, preparation of derivative works, distribution, and display."

20   *Ryan v. Editions Ltd. W., Inc.*, 786 F.3d 754, 760 (9th Cir. 2015) (alteration in original) (quoting

21   *Altera*, 424 F.3d at 1089). "A state law tort claim concerning the unauthorized use of the

22   software's end-product is not within the rights protected by the federal Copyright Act." *Altera*,

23   424 F.3d at 1079. However, in their complaint, Plaintiffs do not allege that Defendants' use of

24   Plaintiffs' code for training purposes violated their rights. Rather, Plaintiffs base their unjust

25

26   [15] Plaintiffs focus this part of their opposition on their breach of open-source license claims, which
     Defendants do not move to dismiss on copyright preemption grounds. ECF No. 66 at 17-18.

27   Plaintiffs additionally argue that their negligence claim and UCL unlawful conduct claim based on
     violation of the DMCA are not preempted. *Id.* Defendants did not move to dismiss Plaintiffs'

28   negligence claims on copyright preemption grounds, and only moved to dismiss Plaintiffs' UCL
     claims to the extent they were based on preempted state law claims.

1   enrichment claim on Defendants' reproduction of Plaintiffs' code as output and Defendants'

2   preparation of derivative works, both of which are rights protected under the Copyright Act.

3   Compl. ¶¶ 204-09 (pleading unjust enrichment claim based on profit derived from both

4   reproduction and preparation of derivative works). Because the rights on which Plaintiffs base

5   their unjust enrichment claim are within those protected by the federal Copyright Act, Plaintiffs'

6   unjust enrichment claim is subject to preemption.

7   Plaintiffs' claim for unjust enrichment is dismissed with leave to amend.

8   **2.   DMCA Section 1202(b) Claim**

9   Because the Court previously dismissed Plaintiffs' claim under Section 1202(a) of the

10   DMCA, the Court now only considers the sufficiency of Plaintiffs' Section 1202(b) claim.

11   "Copyright law restricts the removal or alteration of copyright management information

12   ("CMI") – information such as the title, the author, the copyright owner, the terms and conditions

13   for use of the work, and other identifying information set forth in a copyright notice or conveyed

14   in connection with the work." *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 671 (9th Cir. 2018).

15   Section 1202(b) of the DMCA provides that one cannot, without authority, (1) "intentionally

16   remove or alter any" CMI, (2) "distribute . . . [CMI] knowing that the [CMI] has been removed or

17   altered," or (3) "distribute . . . copies of works . . . knowing that [CMI] has been removed or

18   altered" while "knowing, or . . . having reasonable grounds to know, that it will induce, enable,

19   facilitate, or conceal" infringement. 17 U.S.C. § 1202(b).

20   Plaintiffs allege that their licensed code contains CMI including copyright notices, titles,

21   authors' names, copyright owners' names, terms and conditions for use of the code, and

22   identifying numbers or symbols. Compl. ¶ 144. Plaintiffs allege that Defendants removed or

23   altered that CMI from licensed code, distributed CMI knowing CMI had been removed or altered,

24   and distributed copies of the code knowing that CMI had been removed or altered, all while

25   knowing and possessing reasonable grounds to know that doing so would induce infringement. *Id.*

26   ¶¶ 148, 153-55, 157.

27   Defendants argue that Plaintiffs have not plausibly alleged claims under Sections

28   1202(b)(1) or (b)(3) because these provisions require some active conduct that removes or alters

1   CMI. Defendants argue that the complaint merely alleges "the passive non-inclusion of CMI" by

2   neutral technology which excerpts code without the accompanying CMI, rather than the active

3   removal of CMI from licensed code. ECF No. 50 at 22. This semantic distinction is not

4   meaningful. Plaintiffs allege that the relevant CMI was affixed to their licensed code and that

5   Defendants were aware that such CMI appeared repeatedly across the data used to train Codex and

6   Copilot. Compl. ¶ 92 (CMI "usually appears just before a given block of code"); *id.* ¶ 94 ("[I]n a

7   blog post, GitHub noted[,] 'In one instance, GitHub Copilot suggested starting an empty file with

8   something it had even seen more than a whopping 700,000 different times during training—that

9   was the GNU General Public License.'"). Defendants subsequently trained these programs to

10 ignore or remove CMI and therefore stop reproducing it. *Id.* ¶¶ 94-95 ("As GitHub explains:

11 'GitHub Copilot *has* changed to require a minimum file content.' . . . Copilot no longer reproduces

12 these types of CMI . . . . It has been altered not to.") (emphasis in original). Defendants knew that

13 these programs reproduced training data as output. *Id.* ¶ 90. Plaintiffs thus plead sufficient facts

14 to support a reasonable inference that Defendants intentionally designed the programs to remove

15 CMI from any licensed code they reproduce as output.[16]

16   Defendants further argue that Plaintiffs do not sufficiently plead scienter. As the Ninth

17 Circuit has explained, "the mental state requirement in Section 1202(b) must have a more specific

18 application than the universal possibility of encouraging infringement." *Stevens*, 899 F.3d at 674.

19 *Stevens* involved software that allegedly removed metadata from image files which the defendant

20 subsequently distributed in violation of the DMCA. The Ninth Circuit affirmed the grant of

21 summary judgment to the defendant because the plaintiffs had not offered any evidence that the

22 removal of CMI would impair their policing of infringement. *Id.* at 675. At summary judgment,

23 "specific allegations as to how identifiable infringements 'will' be affected are necessary"; a

24 plaintiff "must make an affirmative showing, such as by demonstrating a past 'pattern of conduct'

25 or 'modus operandi,' that the defendant was aware or had reasonable grounds to be aware of the

26 probable future impact of its actions." *Id.* at 674.

27 _____

28 [16] In other words, Plaintiffs allege that Defendants designed "neutral" programs to effectuate the "non-inclusion" of CMI by actively removing it from Plaintiffs' licensed code.

United States District Court
Northern District of California

1   On a motion to dismiss, a plaintiff must allege sufficient facts to support the reasonable

2 inference that the defendant "knew or had a reasonable basis to know that the removal or alteration

3 of CMI . . . w[ould] aid infringement." *Harrison v. Pinterest, Inc.*, No. 20-cv-05290-EJD, 2022

4 WL 4348460, at * 5 (N.D. Cal. Sept. 19, 2022). At the pleading stage, mental conditions

5 generally need not be alleged with specificity. "Federal Rule of Civil Procedure 9(b) provides that

6 'intent, knowledge, and other conditions of a person's mind may be alleged generally.' Language

7 in *Stevens* . . . does not indicate otherwise; there, the Ninth Circuit held that 'a plaintiff bringing a

8 Section 1202(b) claim must make an affirmative showing' of scienter in the *summary judgment*

9 context." *Logan v. Meta Platforms, Inc.*, --- F. Supp. 3d ---, 2022 WL 14813836, at *9 (N.D. Cal.

10 Oct. 25, 2022) (emphasis in original) (internal citation omitted) (quoting *Stevens*, 899 F.3d at

11 674); *see also Izmo, Inc. v. Roadster, Inc.*, No. 18-cv-06092-NC, 2019 WL 13210561, at *4 (N.D.

12 Cal. Mar. 26, 2019) ("Whether [the defendant] knew or should have known that its activities

13 would induce or enable an infringement of [the plaintiff's] rights is more suited to summary

14 judgment.").

15   Plaintiffs allege that Defendants knew the code they used as training data for Codex and

16 Copilot routinely contained CMI. Compl. ¶ 94. Plaintiffs also allege that GitHub knew that CMI

17 was important for protecting copyright interests. GitHub regularly processed DMCA takedowns,

18 such that it was aware its platform was used to distribute code with removed or altered CMI in a

19 manner which induced infringement. *Id.* ¶ 165-67. Plaintiffs' allegations raise the reasonable

20 inference that Defendants knew or had reasonable grounds to know that removal of CMI carried a

21 substantial risk of inducing infringement. *See, e.g.*, *Schneider v. Youtube, LLC*, No. 20-cv-04423-

22 JD, 2022 WL 3031212, at *2 (N.D. Cal. Aug. 1, 2022) (finding plaintiffs sufficiently pleaded

23 scienter where they alleged that the defendant knew files "routinely contain CMI, that CMI is

24 valuable for protecting copyright holders, and that the distribution of works with missing CMI on

25 [defendant's platform] has induced . . . infringement," supporting a "plausible inference" that

26 defendant removed CMI "with knowledge that doing so carried a 'substantial risk' of inducing

27 infringement") (quoting *Stevens*, 899 F.3d at 676); *Batra v. PopSugar*, No. 18-cv-03752-HSG,

28 2019 WL 482492, at *2 (N.D. Cal. Feb. 7, 2019) (finding plaintiffs' allegations regarding removal

20

1   of CMI supported a "plausible inference" that the defendant "kn[ew] that removing the CMI

2   would help to conceal the alleged infringement").

3   Defendants further argue that Plaintiffs fail to plausibly allege that Defendants distribute

4   CMI "knowing the [CMI] has been removed or altered" in violation of Section 1202(b)(2).

5   Plaintiffs' opposition does not directly address this argument. In the complaint, Plaintiffs allege

6   that "Defendants have a business practice of asserting and/or implying that Copilot is the author of

7   the Licensed Materials" and that "Defendants' false description of the source of Copilot's Output

8   facilitated or concealed infringement by Defendants and Copilot users." Compl. ¶¶ 158-59.

9   Plaintiffs do not identity the assertions, implications, and/or false descriptions of authorship or

10   source at issue, nor do they plead facts that suggest such unidentified statements could constitute

11   CMI. Plaintiffs separately allege that Copilot previously "would sometimes produce [CMI] . . .

12   [which] was not always accurate," but that Copilot "no longer reproduces these types of CMI,

13   incorrect or otherwise, on a regular basis." *Id.* ¶ 95. Plaintiffs plead no specific facts regarding

14   the allegedly inaccurate CMI Copilot once produced alongside output, nor do they plead facts

15   suggesting such inaccurate CMI is likely to be produced alongside their output in the future.

16   Because the allegations in the complaint do not sufficiently allege the distribution of altered CMI,

17   the Court finds that Plaintiffs have failed to state a claim under Section 1202(b)(2).

18   The Court denies Defendants' motions to dismiss Plaintiffs' claim under Sections

19   1202(b)(1) and 1202(b)(3). Plaintiffs' claim under Section 1202(b)(2) is dismissed with leave to

20   amend.

3.   Breach of License Claim

22   Defendants argue that Plaintiffs fail to plead the existence of a contract because they do not

23   indicate which licenses are at issue or which provisions Defendants allegedly breached.[17]

24   Under California law, breach of contract requires plaintiffs to "plead 'the contract,

25   plaintiff's performance (or excuse for nonperformance), defendant's breach, and damage to

26

27   [17] Defendants additionally argue that, to the extent Plaintiffs' breach of license claim is based on
     Defendants' training of Codex and Copilot using licensed code, it is foreclosed by GitHub's
28   Terms of Service. Because Plaintiffs' breach of license claim is not based on such training, *see*
     Compl. ¶¶ 180-83, this argument is irrelevant.

United States District Court
Northern District of California

1   plaintiff therefrom." *Low v. LinkedIn Corp.*, 900 F. Supp. 2d 1010, 1028 (N.D. Cal. 2012)

2   (quoting *Gautier v. Gen. Tel. Co.*, 234 Cal. App. 2d 302, 305 (1965)). "Identifying the specific

3   provision of the contract allegedly breached by the defendant does not require the plaintiff to

4   attach the contract or recite the contract's terms verbatim. Rather, the plaintiff must identify with

5   specificity the contractual obligations allegedly breached by the defendant." *Williams v. Apple,*

6   *Inc.*, 449 F. Supp. 3d 892, 908 (N.D. Cal. 2020) (quoting *Kaar v. Wells Fargo Bank, N.A.*, No. C

7   16-01290 WHA, 2016 WL 3068396, at *1 (N.D. Cal. June 1, 2016)).

8   Plaintiffs advance claims for breach of the eleven suggested licenses GitHub presents to

9   users that require (1) attribution to the owner, (2) inclusion of a copyright notice, and (3) inclusion

10   of the license terms. Compl. ¶ 34 n.4. Plaintiffs attach each of these licenses to the complaint.

11   Plaintiffs allege that use of licensed code "is allowed only pursuant to the terms of the applicable

12   Suggested License," and that each such license requires that any derivative work or copy include

13   attribution, a copyright notice, and the license terms. *Id.* ¶¶ 173, 34 n.4. Plaintiffs further allege

14   that Codex and Copilot reproduce licensed code as output without attribution, copyright notice, or

15   license terms, thereby violating the relevant provisions of each license. While Plaintiffs do not

16   identify the specific subsections of each suggested license that correspond to each of these

17   requirements, the Court finds that Plaintiffs have sufficiently identified "the contractual

18   obligations allegedly breached," as required to plead a breach of contract claim. *Williams*, 449 F.

19   Supp. 3d at 908.

20   Defendants' motions to dismiss Plaintiffs' claim for breach of license is denied.

21   **4.   Unfair Competition**

22   Plaintiffs assert claims for unfair competition under the Lanham Act, the UCL, and

23   California common law against GitHub and OpenAI Defendants. These claims are predicated on

24   Plaintiffs' claims for violation of the DMCA, tortious inference, false designation of origin,

25   violation of the CCPA, and negligence.

26   Defendants move to dismiss Plaintiffs' UCL claims for failure to sufficiently allege

27   predicate claims. To the extent the predicate claims have been dismissed, Plaintiffs' derivative

28   UCL claims must also dismissed with leave to amend. *See Eidmann v. Walgreen Co.*, 522 F.

22

1   Supp. 3d 634, 647 (N.D. Cal. 2021) ("If the 'plaintiff cannot state a claim under the predicate law .

2   . . [the UCL] claim also fails.'") (alterations in original) (quoting *Hadley v. Kellogg Sales Co.*, 243

3   F. Supp. 3d 1074, 1094 (N.D. Cal. 2017)). Plaintiffs' UCL claims predicated upon violation of

4   Sections 1202(a) and 1202(b)(2) of the DMCA, tortious interference, false designation of origin,

5   violation of the CCPA, and negligence are dismissed with leave to amend.

6   Plaintiffs' UCL claims predicated on violation of Sections 1202(b)(1) and (b)(3) of the

7   DMCA remain. Defendants argue that Plaintiffs fail to plead any economic injury arising from

8   the predicate violation, as required for statutory standing under the UCL. To plead a cause of

9   action under the UCL, a plaintiff must allege that the challenged conduct caused "some form of

10   economic injury," like "lost money or property." *Kwikset Corp. v. Super. Ct.*, 51 Cal. 4th 310,

11   323 (2011). As the California Supreme Court explained, a plaintiff may suffer economic injury by

12   "hav[ing] a present or future property interest diminished" or "be[ing] deprived of money or

13   property to which he or she has a cognizable claim." *Id.*

14   Plaintiffs allege that they "have suffered monetary damages" as a result of all of the

15   predicate violations listed in their UCL claim. Compl. ¶ 213. From the allegations in the

16   complaint, however, the Court cannot discern how Defendants' alleged violations of the DMCA

17   have caused or will cause Plaintiffs economic injury. Plaintiffs' opposition argues they "lost the

18   value of [their] work, including their ability to receive compensation as well as the likelihood they

19   would be retained or hired in the future," and have suffered injury to their intellectual property

20   rights, including "loss of value of the computer code and the value to code authors of their

21   attribution rights." ECF No. 66 at 25. However, no such injury is alleged in the complaint.

22   Accordingly, Plaintiffs' UCL claim predicated on violation of the DMCA is dismissed with leave

23   to amend.

24   The OpenAI Defendants also move to dismiss Plaintiffs' claims for unfair competition

25   under the Lanham Act and California common law. Plaintiffs do not address these arguments in

26   opposition, and therefore abandon the claims. *See Diamond S.J. Enter., Inc. v. City of San Jose*,

27   No. 18-cv-01353-LHK, 2018 WL 5619746, at *4 (N.D. Cal. Oct. 29, 2018) ("An opposition

28   brief's failure to address a motion to dismiss'[s] challenges to a claim constitutes abandonment of

23

1   that claim."). Plaintiffs' claims for unfair competition under the Lanham Act and California

2   common law are dismissed with leave to amend.

3   **5.   Civil Conspiracy**

4   Defendants argue that Plaintiffs' civil conspiracy claim must be dismissed because civil

5   conspiracy is not a standalone cause of action.

6   "Conspiracy is not a cause of action, but a legal doctrine that imposes liability on persons

7   who, although not actually committing a tort themselves, share with the immediate tortfeasors a

8   common plan or design in its perpetration." *Applied Equip. Corp. v. Litton Saudi Arabia Ltd.*, 7

9   Cal. 4th 503, 510-11 (1994) (in bank); *see also AccuImage Diagnostics Corp v. Terarecon, Inc.*,

10   260 F. Supp. 2d 941, 947-48 (N.D. Cal. 2003) (dismissing standalone civil conspiracy claim with

11   prejudice because "civil conspiracy is not a separate and distinct cause of action under California

12   law" and explaining that any "amended [civil conspiracy] allegations . . . must be made within the

13   sections of the complaint that contain plaintiff's claims for the underlying" violations).

14   Because Plaintiffs cannot plead civil conspiracy as an independent cause of action, this

15   claim must be dismissed with prejudice.

16   **6.   Declaratory Relief**

17   Defendants move to dismiss Plaintiffs' claim for declaratory relief, arguing that declaratory

18   relief is not a standalone cause of action. Plaintiffs do not address Defendants' arguments as to

19   this claim in their opposition briefs, and therefore have abandoned the claim. *See Diamond S.J.*,

20   2018 WL 5619746, at *4.

21   "[D]eclaratory relief is not a standalone claim." *Mayen v. Bank of Am. N.A.*, No. 14-cv-

22   03757-JST, 2015 WL 179541, at *5 (N.D. Cal. Jan. 14, 2015); *see also Sowinski v. Wells Fargo*

23   *Bank, N.A.*, No. 11-6431-SC, 2012 WL 5904711, at *1 (N.D. Cal. Nov. 26, 2012) (dismissing

24   declaratory relief claim with prejudice and noting that "Plaintiff may still seek declaratory . . .

25   relief in any further pleading, provided that he asserts a claim that could give rise to such relief").

26   Because declaratory relief is not a claim, granting leave to amend would be futile. Accordingly,

27   Plaintiffs' claim for declaratory relief is dismissed with prejudice.

28

CONCLUSION

Defendants' motions to dismiss are granted in part and denied in part. Plaintiffs' claims for violation of Sections 1202(a) and 1202(b)(2) of the DMCA, tortious interference in a contractual relationship, fraud, false designation of origin, unjust enrichment, unfair competition, breach of the GitHub Privacy Policy and Terms of Service, violation of the CCPA, and negligence are dismissed with leave to amend. Plaintiffs' claims for civil conspiracy and declaratory relief are dismissed with prejudice.

Plaintiffs shall file an amended complaint within 28 days of this order.

IT IS SO ORDERED.

Dated: May 11, 2023

_____
JON S. TIGAR
United States District Judge

EXHIBIT F

Joseph R. Saveri (State Bar No. 130064)
Steven N. Williams (State Bar No. 175489)
Cadio Zirpoli (State Bar No. 179108)
Christopher K.L. Young (State Bar No. 318371)
Louis A. Kessler (State Bar No. 243703)
Elissa A. Buchanan (State Bar No. 249996)
Travis Manfredi (State Bar No. 281779)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1000
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
                swilliams@saverilawfirm.com
                czirpoli@saverilawfirm.com
                cyoung@saverilawfirm.com
                lkessler@saverilawfirm.com
                eabuchanan@saverilawfirm.com
                tmanfredi@saverilawfirm.com

Counsel for Individual and Representative
Plaintiffs and the Proposed Class

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

OAKLAND DIVISION

J. DOE 1, et al.,
 	
        Individual and Representative Plaintiffs,
 	
    v.
 	
GITHUB, INC., et al.,
 	
                    Defendants.
	Case Nos. 4:22-cv-06823-JST
	                4:22-cv-07074-JST
	
	PLAINTIFFS' ADMINISTRATIVE MOTION TO FILE UNDER SEAL PORTIONS OF FIRST AMENDED COMPLAINT

I.   INTRODUCTION

Pursuant to Civil Local Rules 7-11 and 79-5, Plaintiffs hereby move the Court to file under seal portions of their First Amended Complaint ("FAC"). The Court previously determined that Plaintiffs may proceed pseudonymously. *See Doe 1 v. GitHub, Inc.*, ___ F. Supp. 3d ___, 2023 WL 3449131, at *7-9 (N.D. Cal. May 11, 2023). The portions sought to be filed under seal contain information that would allow members of the public to discern the identities of Plaintiffs. Accompanying this Motion is the Declaration of Travis Manfredi in Support of Plaintiffs' Administrative Motion to File Under Seal ("Manfredi Decl.") and a proposed order.

II.   ARGUMENT

In the Ninth Circuit, when a party seeks to seal portions of a complaint, the compelling reasons standard is typically used. *See*, *e.g.*, *In re NVIDIA Corp. Derivative Litig.*, C 06-06110 SBA, 2008 WL 1859067, at *3 (N.D. Cal. Apr. 23, 2008); *Ojmar US, LLC v. Sec. People, Inc.*, No. 16-CV-04948-HSG, 2016 WL 6091543, at *2 (N.D. Cal. Oct. 19, 2016); *see also In re ZF-TRW Airbag Control Units Prod. Liab. Litig.*, No. ML1902905JAKFFMX, 2020 WL 13688234, at *3 (C.D. Cal. Sept. 7, 2020) ("Although the Ninth Circuit appears not to have explicitly stated what standard applies to the sealing of a complaint, many courts in this district and elsewhere have found that the compelling reasons standard applies."). This is because complaint is "the root, the foundation, the basis by which a suit arises and must be disposed of." *NVIDIA Corp. Derivative Litig.*, 2008 WL 1859067, at *3.

Compelling reasons "outweigh the public's interest in disclosure and justify sealing court records . . . when such 'court files might have become a vehicle for improper purposes,' such as the use of records to gratify private spite, promote public scandal, circulate libelous statements, or release trade secrets . . . ." *Kamakana v. City & Cnty. of Honolulu*, 447 F.3d 1172, 1179 (9th Cir. 2006) (citing *Nixon v. Warner Commc'ns, Inc.*, 435 U.S. 589, 598 (1978)). The party seeking to seal bears the burden of showing compelling reasons apply. *See id.* at 1178.

This Court has already determined Plaintiffs may proceed pseudonymously at this stage of the case under the five-factor test applicable in the Ninth Circuit. *Doe 1*, 2023 WL 3449131, at *7–9 (citing, *inter alia*, *Doe v. Kamehameha Schs./Bernice Pauahi Bishop Est.*, 596 F.3d 1036, 1042

1 (9th Cir. 2010) & *Does I Thru XXIII v. Advanced Textile Corp.*, 214 F.3d 1058, 1068 (9th Cir.

2 2000)). Amongst other things, the Court determined that proceeding pseudonymously at this

3 stage is appropriate because Plaintiffs harbored reasonable fear that threats to their lives may be

4 carried out against them if their identities were to become public. *Id.* at *8.[1] This conclusion is

5 equally applicable here where distinctive descriptors of Plaintiffs' unique code would enable

6 members of the public to discern Plaintiffs' true identities. Given Plaintiffs' reasonable fear of

7 physical reprisal, Plaintiffs should be permitted to file under seal information that could lead to

8 the disclosure of their true identities.

9   III.   CONCLUSION

10   For the reasons stated above, the Plaintiffs request the following paragraphs of the FAC

11 be filed under seal:

FAC Paragraph(s) #	Description of Information	Reason for Sealing
19–23	Each paragraph contains one Plaintiff's name	The Court has already determined it is appropriate for Plaintiffs to proceed pseudonymously at this stage due to Plaintiffs' reasonable fear that credible threats of physical violence may be carried out. *See Doe 1*, 2023 WL 3449131, at *7-9.
101–103	Examples and descriptions of code Doe 2 made available to the public on GitHub.	To protect Doe 2 from credible threats of physical violence. Doe 2 could easily be identified by searching for this code on GitHub and other public repositories.
106–109 & 111	Examples and descriptions of code Doe 1 made available to the public on GitHub.	To protect Doe 1 from credible threats of physical violence. Doe 1 could easily be identified by searching for this code on GitHub and other public repositories.

---

[1] The Court also determined that Defendants are not prejudiced because Plaintiffs have already disclosed their true names to them subject to the underlying protective order, and that the public interest factor does not weigh against anonymity at this stage. *Id.* at *8–9.

ADMINISTRATIVE MOTION TO FILE UNDER SEAL PORTIONS OF FIRST AMENDED COMPLAINT

114–118 & 122–125	Examples and descriptions of code Doe 5 made available to the public on GitHub.	To protect Doe 5 from credible threats of physical violence. Doe 5 could easily be identified by searching for this code on GitHub and other public repositories.

Manfredi Decl. ¶ 4.

Dated: June 8, 2023

By: _/s/ Joseph R. Saveri_
Joseph R. Saveri

Joseph R. Saveri (State Bar No. 130064)
Steven N. Williams (State Bar No. 175489)
Cadio Zirpoli (State Bar No. 179108)
Christopher K.L. Young (State Bar No. 318371)
Louis A. Kessler (State Bar No. 243703)
Elissa A. Buchanan (State Bar No. 249996)
Travis Manfredi (State Bar No. 281779)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1000
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
                         swilliams@saverilawfirm.com
                         czirpoli@saverilawfirm.com
                         cyoung@saverilawfirm.com
                         lkessler@saverilawfirm.com
                         eabuchanan@saverilawfirm.com
                         tmanfredi@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
Telephone: (323) 968-2632
Facsimile: (415) 395-9940
Email: mb@butericklaw.com

Counsel for Individual and Representative Plaintiffs and the Proposed Class

ADMINISTRATIVE MOTION TO FILE UNDER SEAL PORTIONS OF FIRST AMENDED COMPLAINT

EXHIBIT G

1 ANNETTE L. HURST (SBN 148738)
   ahurst@orrick.com
2 DANIEL D. JUSTICE (SBN 291907)
   djustice@orrick.com
3 ORRICK, HERRINGTON & SUTCLIFFE LLP
   405 Howard Street
4 San Francisco, CA 94105-2669
   Telephone: +1 415 773 5700
5 Facsimile: +1 415 773 5759

6 WILLIAM W. OXLEY (SBN 136793)
   woxley@orrick.com
7 ALYSSA CARIDIS (SBN 260103)
   acaridis@orrick.com
8 ORRICK, HERRINGTON & SUTCLIFFE LLP
   355 S. Grand Avenue
9 Los Angeles, CA 90071
   Telephone: +1 213 629 2020
10 Facsimile: +1 213 612 2499

11 *Attorneys for GitHub, Inc. and Microsoft Corporation*

12   UNITED STATES DISTRICT COURT

13   NORTHERN DISTRICT OF CALIFORNIA

14   OAKLAND DIVISION

15

16 J. DOE 1, et al.,   Case No. 4:22-cv-6823-JST

17   Individual and   Consolidated with Case No. 4:22-cv-7074-JST
            Representative Plaintiffs,
18   **DEFENDANTS GITHUB AND**
        v.   **MICROSOFT'S NOTICE OF MOTIONS**
19   **AND MOTIONS TO DISMISS PORTIONS**
   GITHUB, INC., et al.,   **OF THE FIRST AMENDED COMPLAINT**
20   **IN CONSOLIDATED ACTIONS**
            Defendants.
21   Date: September 14, 2023
                                              Time: 2:00 p.m.
22 AND CONSOLIDATED ACTION   Courtroom: 6, 2d Floor
                                              Judge: Hon. Jon S. Tigar
23
                                              Amended Complaint Filed: June 8, 2023
24

25

26

27

28

TABLE OF CONTENTS

i

GitHub and Microsoft's MTD Am. Compl.
No. 4:22-cv-6823-JST

TABLE OF AUTHORITIES

Page(s)

Cases

1

NOTICE OF MOTIONS AND MOTIONS

2

TO PLAINTIFFS AND THEIR ATTORNEYS OF RECORD:

3

PLEASE TAKE NOTICE THAT on September 14, 2023 at 2:00 p.m., before the

4

Honorable Jon S. Tigar seated in Courtroom 6 of the United States Courthouse at Oakland,

5

California, with appearances to be made by Zoom videoconference unless otherwise ordered by

6

the Court, Defendants GitHub, Inc. ("GitHub") and Microsoft Corporation ("Microsoft") will,

7

and hereby do, move, pursuant to Federal Rules of Civil Procedure, 12(b)(1) and 12(b)(6) to

8

dismiss portions of the operative Amended Complaint in these consolidated actions, filed as ECF

9

No. 98 in the 4:22-cv-6823 action (the "Amended Complaint" or "Am. Compl."), as to both

10

GitHub and Microsoft.

11

The grounds for the Motions are as follows. First, insofar as Plaintiffs attempt to plead

12

any claims for damages in the Amended Complaint, those claims should be dismissed pursuant to

13

Rule 12(b)(1) because Plaintiffs fail to allege that they suffered any actual injury at the hands of

14

either GitHub or Microsoft. Accordingly, all requests for or assertions of entitlement to monetary

15

relief should be dismissed from the case, and Plaintiffs' requests for punitive damages should be

16

independently dismissed pursuant to Rule 12(b)(6) because Plaintiffs fail to plead entitlement to

17

such relief. Second, Counts 1, 4, 5, 6, 7, and 8 of the Amended Complaint should be dismissed

18

pursuant to Rule 12(b)(6) because they are either (a) legally foreclosed, (b) deficient because

19

factual allegations required to support necessary elements are missing or implausible, or (c) both.

20

21

22

23

24

25

26

27

28

GITHUB AND MICROSOFT'S MTD AM. COMPL.
NO. 4:22-CV-6823-JST

1

MEMORANDUM OF POINTS AND AUTHORITIES

2

INTRODUCTION AND SUMMARY OF ISSUES

3   The central premise of Plaintiffs' case is that GitHub's Copilot AI tool, in response to user

4   input, is capable of generating coding suggestions that are reproductions of Plaintiffs' code. But

5   Plaintiffs' initial Complaint had a critical defect. As this Court found, Plaintiffs failed to allege

6   that Copilot had ever actually generated any suggestion reproducing *their* code, leaving Plaintiffs

7   uninjured and therefore without standing to pursue damages. Lacking real-life instances of harm,

8   Plaintiffs now try to manufacture some. In their Amended Complaint, Plaintiffs set about to

9   create "a few examples of Plaintiffs' code being reproduced," Am. Compl. ¶ 127, by themselves

10 mechanistically entering their own code into a code editor until Copilot generated a suggested

11 completion that looked like (but still differed from) Plaintiffs'. Far from fixing the defect,

12 however, Plaintiffs' contrived examples simply confirm how implausible it is that Copilot has

13 ever actually generated a copy of Plaintiffs' code.

14   The problem with Plaintiffs' examples is that there is nothing in the Amended Complaint

15 to suggest that the prompts Plaintiffs employed would ever actually happen in the real world. In

16 one example, Plaintiffs precisely transcribed 22 lines of a function that sets up a game board—

17 including the name of the game being set up, layout of the board, and number and type of

18 pieces—before Copilot suggested an 18-line completion that was not even an exact copy (more

19 on that in a moment). It is inconceivable that an actual user would somehow precisely type in 22

20 lines of that code by chance or accident while working on her own coding project. That is why

21 Plaintiffs do not even attempt to allege that any real-life user would or could come anywhere

22 close to entering prompts anything like those that yielded Plaintiffs' examples. The fact that

23 Plaintiffs had to go to such unrealistic lengths to manufacture even the few examples they include

24 in the Amended Complaint only reaffirms how unlikely it is that Plaintiffs have actually suffered

25 any injury. Plaintiffs' attempt to manufacture standing fails. *Infra* § I.

26   Plaintiffs' examples also undermine their theory on the merits—and here it is not the

27 prompts, but the outputs that are the problem. Plaintiffs continue to allege that Copilot cannot

28 really code "the way a human would," Am. Compl. ¶ 58, and instead is just "reproducing," Am.

GITHUB AND MICROSOFT'S MTD AM. COMPL.
NO. 4:22-CV-6823-JST

1   Compl. ¶ 77. But Plaintiffs were unable to manipulate Copilot into emitting a "verbatim copy of

2   copyrighted code." Am. Compl. ¶ 78. Their examples all contain variations from their code, just

3   as you might expect from an AI model that has learned common functional concepts across a

4   universe of code, then generated context-dependent suggestions that incorporate that learning.

5   Indeed, one example Plaintiffs tout is nothing more than a list of known chemical compounds in

6   the same order as a list widely available on the internet—and even that is not a verbatim

7   reproduction of Plaintiffs' version. This is all of little surprise, since Plaintiffs elsewhere admit

8   that the way Copilot actually works is by discerning "statistically significant patterns." Am.

9   Compl. ¶ 85. Because Plaintiffs do not allege that any of *their* code is prevalent enough to be

10 detected as a pattern in the training set, it is implausible that Copilot would ever output that

11 code—even assuming, as necessary at this stage, that "about 1% of the time" Copilot may more

12 precisely reproduce a snippet of *someone else's* oft-repeated code, Am. Compl. ¶ 94.

13   That is why Plaintiffs are now forced to concede that, even in the tiny universe of

14 contrived Copilot outputs they created, "more often" those outputs *differ* from their code. Am.

15 Compl. ¶ 96. And to accommodate this reality, the Amended Complaint vastly expands their

16 legal theory to reach not just actual reproductions, but also "functionally equivalent" code

17 suggestions. Am. Compl. ¶ 103. But it is beyond dispute that copyright protection does not

18 extend to the functionality (as opposed to the expression) of the underlying code. *See Sony*

19 *Comput. Ent., Inc. v. Connectix Corp.*, 203 F.3d 596, 599 (9th Cir. 2000); *Apple Comput., Inc. v.*

20 *Microsoft Corp.*, 35 F.3d 1435, 1444-45 (9th Cir. 1994); *cf. Google LLC v. Oracle Am., Inc.*, 141

21 S. Ct. 1183, 1197-1200 (2021). So Plaintiffs' new reliance on variations with alleged functional

22 equivalency does not work for their claims under § 1202(b). While GitHub and Microsoft do not

23 re-raise here the arguments this Court previously addressed with respect to Plaintiffs' § 1202(b)

24 claims, this Court did not expressly resolve whether Plaintiffs' failure to allege that CMI was

25 removed from identical copies meets the statutory requirement that the resulting infringement

26 concern "copies … of a work," a requirement Plaintiffs' new theory could never satisfy. The

27 Amended Complaint's new reliance on algorithmic similarities underscores the need for this

28 Court to reach the "copies" limitation of § 1202(b). *Infra* § II.

1   Finally, Plaintiffs' Amended Complaint includes one more expansion worthy of note.

2   Previously, the Court was skeptical whether the original Complaint was meant to premise liability

3   merely on the training of machine learning models, rather than on outputs. Plaintiffs now

4   advance two claims—unjust enrichment and negligence—that clearly attack training alone as

5   purportedly unlawful: "Defendants have unjustly utilized access to Licensed Materials hosted on

6   GitHub to create Codex and Copilot." Am. Compl. ¶ 269. These claims, as well as all of

7   Plaintiffs' other tort claims are still defective. Each is based on either the alleged reproduction or

8   distribution of code, or use of code to prepare a derivative work, and is therefore preempted by

9   § 301 of the Copyright Act. *Infra* § III. And each has additional defects as well. *Infra* § IV.

10   **FACTUAL AND PROCEDURAL BACKGROUND**

11   **A.   The Technology At Issue – According To The Amended Complaint.**

12   **1.   OpenAI Develops A Generative AI Tool Called Codex.**

13   OpenAI is a nonprofit organization that develops machine learning models, also referred

14   to as "Artificial Intelligence." Am. Compl. ¶¶ 2, 167, 169. Such models are typically trained

15   through exposure to a corpus of material called "training data." Am. Compl. ¶ 85. The patterns

16   discerned from the set of training data become part of the model, which can then generate

17   answers based upon those patterns in response to user prompts. Am. Compl. ¶ 85; *see* Am.

18   Compl. ¶ 56.

19   The model at issue in this case is called Codex. Codex is a generative AI model trained

20   on publicly available computer source code. Am. Compl. ¶¶ 56, 176. The model embodies

21   "inferred … statistical patterns governing the structure of code," Am. Compl. ¶ 56, which it has

22   discerned from the training data based on "a complex probabilistic process." Am. Compl. ¶ 83.

23   It is thus capable, in response to a prompt, of "predic[ting] … the most likely [coding] solution."

24   Am. Compl. ¶ 83. "[E]ssentially [it] returns the solution it has found in the most [coding]

25   projects when those projects are somehow weighted to adjust for whatever variables [the model]

26   ha[s] identified as relevant." Am. Compl. ¶ 83.

27   Generative AI models are capable of "simulat[ing] human reasoning or inference,"

28   engaging in the same sort of pattern recognition, synthesis, and prediction we do. Am. Compl.

¶ 2. AI models like Codex also enable statistical analysis and prediction vastly more powerful, efficient, and sensitive than what the human brain can accomplish. Am. Compl. ¶ 85. At the same time, Codex "does not understand the meaning of code," Am. Compl. ¶ 58, nor its "semantics and context the way humans do," Am. Compl. ¶ 85. Codex thus offers both the ingenious and the mundane, a powerful tool of invention for humans who supply the insight to direct its range of performance.

2.   GitHub Offers Copilot, A Code Completion Tool Based On Generative AI.

GitHub Copilot is a programming assistant. Am. Compl. ¶ 8. The Amended Complaint alleges that Copilot "uses the OpenAI Codex to suggest code and entire functions in real-time" to software developers. Am. Compl. ¶ 51. To use Copilot, a GitHub user installs it "as an extension to various code editors, including Microsoft's Visual Studio and VS Code." Am. Compl. ¶ 71. "As the user types [code] into the editor," Copilot treats the user's input as a prompt, generating suggestions for code that may be appropriate for the developer's purposes. Am. Compl. ¶ 71. Copilot is a subscription tool available to GitHub users for $10 per month or $100 per year. Am. Compl. ¶ 8.

Plaintiffs allege that the generative AI model that powers Copilot was trained on billions of lines of code that GitHub users stored in public GitHub repositories. *See* Am. Compl. ¶¶ 87, 186. When GitHub users put their code on GitHub, they choose whether to make the code repositories private or public. Am. Compl. ¶ 160. Users who set their repositories "to be viewed publicly … grant each User of GitHub a nonexclusive, worldwide license to use, display, and perform Your Content through the GitHub Service and to reproduce Your Content solely on GitHub as permitted through GitHub's functionality." Am. Compl. Ex. 1 at 27 (GitHub Terms of Service ("TOS") at 7). Every user agrees to GitHub's TOS, which include a "License Grant" to GitHub to "store, archive, parse, and display … and make incidental copies" as well as "parse it into a search index or otherwise analyze it" and "share" the content in public repositories with other users. Am. Compl. Ex. 1 at 26-27 (GitHub TOS at 6-7). And users can also select from a range of preset open source licenses to apply to the code published in their various GitHub repositories, apply their own individual licenses, or select none at all. Am. Compl. ¶ 38 n.4 &

Appx. A.

Any GitHub user thus appreciates that code placed in a public repository is genuinely public. Anyone is free to examine, learn from, and understand that code, as well as repurpose it in various ways. And, consistent with this open source ethic, neither GitHub's TOS nor any of the common open source licenses prohibit either humans or computers from reading and learning from publicly available code. *See* Am. Compl. ¶ 38 n.4 & Appx. A.

B.   Prior Proceedings.

In the initial Complaint, Plaintiffs raised twelve claims against Defendants. Compl., Doe 3 v. GitHub, Inc., No. 22-cv-7074-JST (N.D. Cal. Nov. 10, 2022), ECF No. 1. GitHub and Microsoft moved to dismiss the entirety of the initial operative complaint pursuant to Federal Rules of Civil Procedure 8, 9, 10, 12(b)(1), and 12(b)(6). Mots. to Dismiss Operative Compl. in Consol. Actions, ECF No. 50.

This Court granted in part and denied in part GitHub and Microsoft's motions to dismiss. ECF No. 95. On standing, this Court held that Plaintiffs "failed to establish an injury-in-fact sufficient to confer standing for their claims for damages based on injury to property rights." *Id.* at 10. It recognized, *id.* at 6, that "at the pleading stage, the plaintiff must 'clearly … allege facts demonstrating each element'" of standing—(1) a personal injury in fact; (2) that the injury was likely caused by the defendant, and (3) that the injury would likely be redressed by judicial relief. *Spokeo, Inc. v. Robins*, 578 U.S. 330, 338 (2016) (quoting *Warth v. Seldin*, 422 U.S. 490, 518 (1975)). And it found that Plaintiffs had failed to demonstrate that "they themselves" had suffered injury from "Copilot reproduc[ing] Plaintiffs' code as output with missing or incorrect attribution, copyright notices, and license terms." Order 8, ECF No. 95. Plaintiffs therefore lacked standing to pursue retrospective relief (*i.e.*, damages). This Court also held that Plaintiffs failed to allege facts demonstrating an injury-in-fact sufficient to confer standing for their privacy-based claims. *Id.* at 7.

Nevertheless, this Court held that "Plaintiffs plausibly allege that there is at least a substantial risk that Defendants' programs will produce Plaintiffs' licensed code as output," and that this risk "may support standing for injunctive relief." *Id.* at 9. The Court noted, however,

1   that "Plaintiffs do not allege they were injured by Defendants' use of licensed code as training

2   data[.]" *Id.* at 8 n.7.

3   On the merits, this Court dismissed Plaintiffs' claims for violation of § 1202(a) of the

4   DMCA, tortious interference, fraud, false designation of origin, and violation of the CCPA for

5   failure to state a claim under Rule 12(b)(6), with leave to amend. *Id.* at 16. It further dismissed

6   the unjust enrichment claim as preempted by the Copyright Act, with leave to amend. *Id.* at 17-

7   18. As to Plaintiffs' § 1202(b) claim, this Court ruled that Plaintiffs pled sufficient facts to

8   support a reasonable inference that Defendants intentionally designed the programs to remove

9   CMI from any licensed code they reproduce as output, *id.* at 19-21, but did not address GitHub

10   and Microsoft's argument that the § 1202(b) claim fails because the Complaint did not allege

11   CMI removal from an "identical copy of the work." *See* Mots. to Dismiss 13-14, ECF No. 50.

12   This Court denied Defendants' motions to dismiss the breach of contract claim, Order at

13   22, ECF No. 95. It then dismissed the UCL claims to the extent the predicate claims had been

14   dismissed, with leave to amend, *id.* at 22-23, and it further dismissed the UCL claims predicated

15   on violations of § 1202(b)(1) and (b)(3) of the DMCA, also with leave to amend, *id.* at 23. The

16   Court dismissed the civil conspiracy and declaratory relief counts with prejudice. *Id.* at 24.

17   **C.   Plaintiffs' New Allegations.**

18   In their Amended Complaint, Plaintiffs have joined a fifth J. Doe GitHub user as a new

19   Plaintiff and alleged eight claims for relief. Am. Compl. ¶¶ 19-23. The allegations underlying

20   Plaintiffs' claims under § 1202(b)(1) and (3), Am. Compl. ¶¶ 183-213 (Count 1), and for breach

21   of open source licenses, Am. Compl. ¶¶ 214-29 (Count 2), are largely unchanged from the initial

22   Complaint. Plaintiffs also now include claims for breach of contract for selling licensed materials

23   in violation of GitHub's policies, Am. Compl. ¶¶ 230-40 (Count 3); intentional interference with

24   prospective economic relations, Am. Compl. ¶¶ 241-53 (Count 4); negligent interference with

25   prospective economic relations, Am. Compl. ¶¶ 254-65 (Count 5); a modified unjust enrichment

26   claim, Am. Compl. ¶¶ 266-74 (Count 6); a modified unfair competition claim, Am. Compl.

27   ¶¶ 275-81 (Count 7); and a new negligence claim, Am. Compl. ¶¶ 282-89 (Count 8).

28   In support of these claims, Plaintiffs continue to allege that they "published Licensed

                                                          6   GITHUB AND MICROSOFT'S MTD AM. COMPL.
                                                                                          NO. 4:22-CV-6823-JST

1   Materials they owned a copyright interest in to at least one GitHub repository under one of the

2   Suggested licenses." Am. Compl. ¶¶ 19-23. Plaintiffs allege that because Copilot was trained on

3   public GitHub repositories, they can be "reasonably certain" their Licensed Materials were

4   "ingested by Copilot and [are] sometimes returned to users as Output." Am. Compl. ¶ 97.

5   New to the Amended Complaint, however, is Plaintiffs' acknowledgement that "more

6   often" than not, a snippet reproducing Plaintiffs code are snippets with "variations" from existing

7   code. Am. Compl. ¶ 96. Three of the Plaintiffs also allege, with respect to code in their

8   repositories, that they were able to "prompt[] Copilot" to "emit[]" their own Licensed Materials.

9   Am. Compl. ¶¶ 98, 100, 103, 110, 123, 125, 127-128.

10   In the first example, Plaintiffs prompted Copilot with a Java *enum* keyword, which is used

11   to represent (enumerate) a list of constants. Am. Compl. § 101. The prompt also contained the

12   name for the *enum* as ▮▮▮▮▮▮ followed by a line of code reflecting the specific format in

13   which the constants would be listed—namely, starting with the name of the ▮▮▮▮▮

14   ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮." Based on this

15   information, Copilot suggested the remaining ▮▮▮▮▮▮▮▮▮ in the same format—

16   in the same order as Plaintiffs' own code, but also in the same order as a list widely available on

17   the internet. *Compare* Am. Compl. ¶¶ 100-04 *with* http://en.wikipedia.org/wiki/

18   ▮▮▮▮▮▮▮▮.

19   The other three examples make clear that Plaintiffs ventured far outside any ordinary use

20   of Copilot to manufacture them, and even then failed to get Copilot to produce an actual copy. In

21   one example, Plaintiffs entered, character for character, 22 lines of Doe 1's code to yield a

22   suggested output of 18 additional lines of code, which Plaintiffs nonetheless concede "is not an

23   exact match for Doe 1's code." Am. Compl. ¶ 108. Plaintiffs characterize these differences as "a

24   modification *based on a copy* of Doe 1's code," Am. Comp. ¶ 108 (emphasis added)—an

25   insinuation with no supporting allegation. But in any event, a comparison of Plaintiffs' code

26   against Copilot's output speaks for itself, demonstrating extensive differences between the two:

27

28

GITHUB AND MICROSOFT'S MTD AM. COMPL.
NO. 4:22-CV-6823-JST
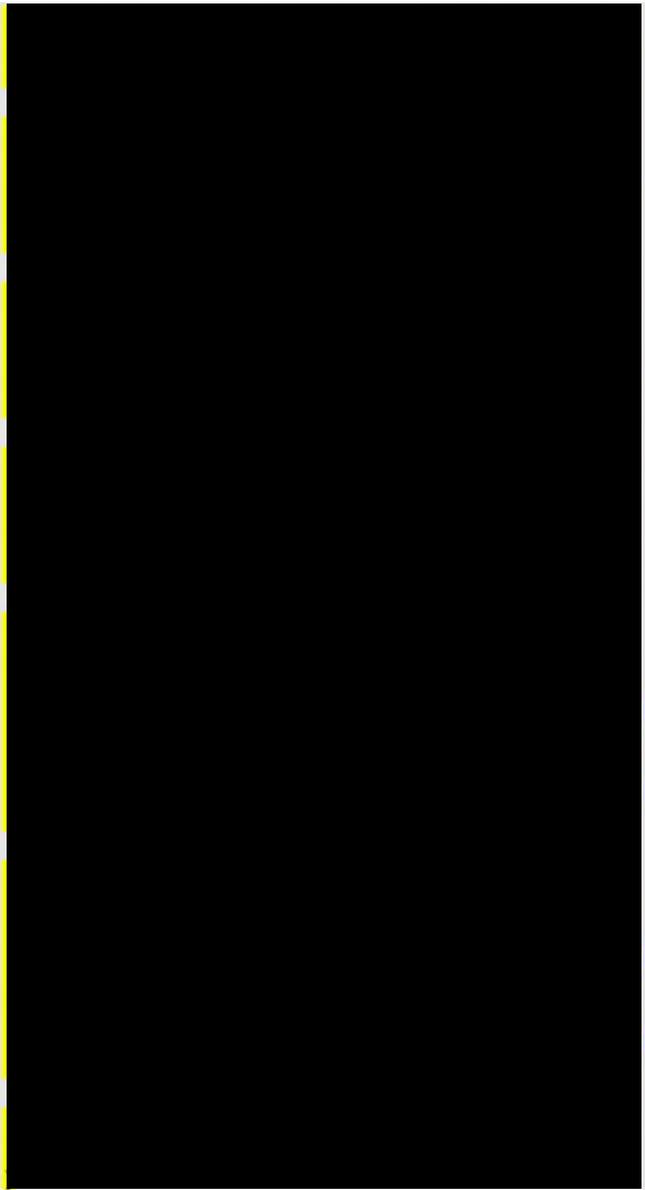
Original code: Output:
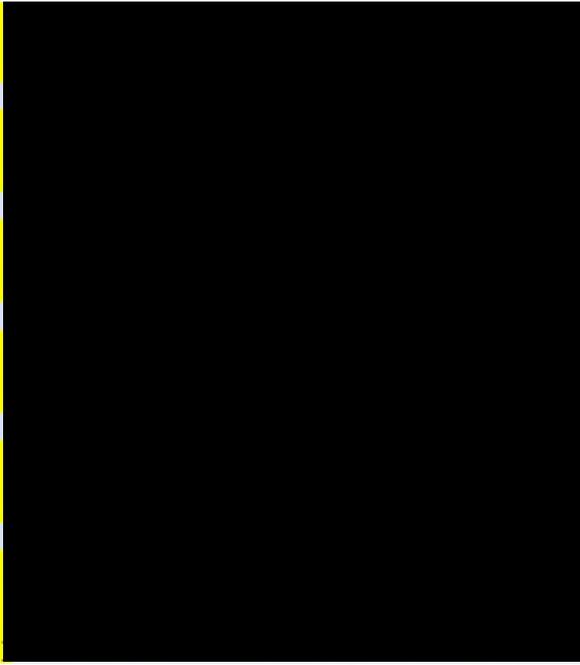
Am. Compl. ¶¶ 106-07.

The same is true of the remaining two examples. Am. Compl. ¶¶ 113-126. Plaintiffs offer

no example of a verbatim copy of code.

Although most of Plaintiffs' claims are focused on the same output-without-attribution

theory Plaintiffs advanced in their initial Complaint, their claims for unjust enrichment and

negligence are based on the training of Copilot itself. *See* Am. Compl. ¶ 269 ("Defendants have

unjustly utilized access to Licensed Materials hosted on GitHub to create Codex and Copilot.");

GITHUB AND MICROSOFT'S MTD AM. COMPL.
NO. 4:22-CV-6823-JST

1   *id.* ¶ 271 ("Plaintiffs did not consent to the unauthorized use of their Licensed Materials to train

2   Codex and Copilot."); *id.* ¶ 284 ("Defendants breached their duties by … engineering, designing,

3   maintaining, and controlling systems—including Codex and Copilot—which are trained on

4   Plaintiffs' and Class members' Licensed materials without their authorization."). Plaintiffs allege

5   that this training constitutes the creation of a "Derivative Work." Am. Compl. ¶ 194 n.34.

6   **ARGUMENT**

7   **I.   THE AMENDED COMPLAINT DOES NOT ESTABLISH ARTICLE III
             STANDING TO PURSUE CLAIMS FOR DAMAGES.**

8

9   In their Amended Complaint, Plaintiffs attempt to manufacture the standing to seek

10   damages that this Court found lacking in the initial Complaint. Plaintiffs added a new Plaintiff

11   who alleges that they were able to generate a copy of their code (after inputting many lines of

12   their own code as a prompt), and added allegations from two other Plaintiffs claiming Copilot

13   outputs that look like their code. Am. Compl. ¶¶ 100-28. Plaintiffs also attempt to predicate their

14   claims for unjust enrichment and negligence solely on the alleged use of code in Plaintiffs'

15   repositories to train Codex and Copilot, seemingly suggesting that training *itself* causes them

16   injury. Am. Compl. ¶¶ 269, 284. Each of these artificial attempts to generate standing to pursue

17   monetary damages fails, requiring dismissal of any request for damages under Rule 12(b)(1).

18   **A.   Plaintiffs Cannot Manufacture Standing To Seek Damages.**

19   ***Self-inflicted "injury" cannot confer standing.*** Plaintiffs are still unable to plead facts

20   plausibly suggesting that Copilot has ever generated a copy of Plaintiffs' code when Copilot is

21   prompted by normal use, rather than to generate litigation-focused outputs. None of their code is

22   alleged to be popular or repeated elsewhere, nor well-suited to any purpose any other user may

23   hope to satisfy when working on a coding project using Copilot. Plaintiffs' theory of injury thus

24   continues to rest on the bare notion that code in their repositories went *into* Copilot via training,

25   so it could conceivably come *out of* Copilot, too.

26   Plaintiffs try to make this bare possibility more concrete by including several examples

27   that they "have been able to generate" by entering prompts into Copilot. Am. Compl. ¶ 128. If

28   Plaintiffs mean to suggest that these outputs themselves have injured them, that is a non-starter.

                                                              9   GITHUB AND MICROSOFT'S MTD AM. COMPL.
                                                                                       NO. 4:22-CV-6823-JST

1   Plaintiffs may not "manufacture standing merely by inflicting harm on themselves." *Clapper v.*

2   *Amnesty Int'l. USA*, 568 U.S. 398, 416 (2013). Such self-inflicted injuries have a traceability

3   problem—"[a] self-inflicted injury, by definition, is not traceable to anyone but the plaintiff."

4   *Buchholz v. Meyer Njus Tanick, PA*, 946 F.3d 855, 866-67 (6th Cir. 2020) (collecting authorities).

5   As one court put it in the context of a claim about misappropriation of the plaintiff's likeness,

6   "[i]t is silly to complain at length that one is devastated by the *unauthorized* use of his yearbook

7   photo to sell a product where *he orchestrated the use* and the injury may not have occurred but

8   for his initiative." *Fry v. Ancestry.com Operations Inc.*, No. 22-CV-140 JD, 2023 WL 2631387,

9   at *5 (N.D. Ind. Mar. 24, 2023) (emphases in original). So it is here.

10   Nor does including these examples in the Amended Complaint raise an inference that

11   Copilot has elsewhere generated copies of code in the course of ordinary use by developers other

12   than Plaintiffs. Quite the opposite: the lengths Plaintiffs and their counsel seemingly had to go to

13   in order to conjure the four examples in the Amended Complaint weigh strongly against such an

14   inference. Take the examples pertaining to Doe 5, the only Plaintiff who has not already been

15   ruled to lack damages standing. The Amended Complaint alleges that by "prompt[ing]" Copilot

16   with the first *several hundred* characters of code *precisely* as it appears in Doe 5's coding project

17   on GitHub, Plaintiffs were able to get Copilot to output code that completes the function that

18   Plaintiffs asked Copilot to write in a way that is similar, but not identical, to how that particular

19   Doe coded it. Am. Compl. ¶¶ 121-26.

20   It is impossible that any ordinary user would prompt Copilot with a precise replica of long

21   portions of Plaintiffs' code without deliberately transcribing it the way Plaintiffs did. Plaintiffs

22   certainly do not allege that any of these prompts could ever actually occur. In any event, it is

23   Plaintiffs' burden to establish standing, *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 561 (1992),

24   and so it is Plaintiffs' job to show that the examples in their Amended Complaint are somehow

25   exemplary of something that actually could have happened in real life. They have still not

26   identified any prompt an ordinary user of Copilot would ever be likely to use that would generate

27   a copy of code in their repositories. They therefore continue to lack standing to pursue damages.

28   ***Standing cannot be predicated on events that post-date the filing of suit.*** For Does 1 and

1   2, Plaintiffs' attempt to newly manufacture standing by including examples fails for another

2   reason: "'[S]tanding is determined as of the date of the filing of the complaint. The party

3   invoking the jurisdiction of the court cannot rely on events that unfolded after the filing of the

4   complaint to establish its standing.'" *Wilbur v. Locke*, 423 F.3d 1101, 1107 (9th Cir. 2005)

5   (alteration adopted), *abrogated on other grounds by Levin v. Commerce Energy, Inc.*, 560 U.S.

6   413 (2010). Plaintiffs were all adjudged to lack standing to pursue damages based on the facts

7   alleged in the initial Complaint. The only additional facts in the Amended Complaint that even

8   conceivably go to Plaintiffs' claimed injury are the artificially manufactured examples, all of

9   which appear to have been created after this Court dismissed Plaintiffs' damages claims. A

10 plaintiff may not belatedly amend the jurisdictional facts to generate standing where none was

11 present at the time the Complaint was filed. *See Newman-Green, Inc. v. Alfonzo-Larrain*, 490

12 U.S. 826, 831 (1989).

13   ***Plaintiffs identify no cognizable injury from training.*** In its ruling on the prior motions

14 to dismiss, this Court noted Plaintiffs' "suggest[ion] [that] they were also injured by Defendant's

15 use of Plaintiffs' licensed code as training data." Order 8 n.7, ECF No. 95. But it found that

16 Plaintiffs had failed to "describe such an injury." *Id.* And indeed, at the hearing on Defendants'

17 motions, this Court repeatedly explained that "[i]f the training had occurred and then nothing else,

18 this case would not be here." Tr. 23. The Court expressed "difficulty in understanding how the

19 training aspect of this product injures anybody's rights," noting that if any person could freely log

20 onto GitHub to view code in public repositories, "why can't a software program do the same

21 thing?" Tr. 27-28. And when the Court asked whether training alone could violate any

22 attribution requirement in any applicable open-source license, Plaintiffs' counsel could only

23 respond "Perhaps it doesn't." Tr. 31.

24   Nevertheless, Plaintiffs now appear to predicate two claims for unjust enrichment (Count

25 6) and Negligence (Count 8) on the act of training alone. These claims are highly dubious on the

26 merits, as discussed below. But even if they were not, Plaintiffs still fail to allege any cognizable

27 injury to them that would result from the mere training of a generative AI model based, in part, on

28 code contained in Plaintiffs' repositories. Plaintiffs do not even try to allege how they have been

11   GITHUB AND MICROSOFT'S MTD AM. COMPL.
                                               NO. 4:22-CV-6823-JST

1 harmed by use of code in their public repositories to train a machine learning model. They do not

2 identify a "concrete, particularized" invasion of a recognized legal interest that is violated merely

3 by a software program viewing and learning from data that is publicly available on the internet.

4 And Plaintiffs suggest no harm that would flow from the fact that an algorithm happens to exist

5 that was in part trained on code Plaintiffs chose to make publicly available to anyone. Standing

6 to seek damages for purposes of Counts 6 and 8 therefore cannot be predicated on training.

**B.   All Requests For Or Assertions Of Entitlement To Monetary Relief Should Be
Dismissed.**

9 Because Plaintiffs lack standing to seek damages, all requests for or assertions of

10 entitlement to such relief should be dismissed from the case. This includes including Plaintiffs'

11 allegations concerning a putative "Damages Class," Am. Compl. at 9:13-16, which no named

12 Plaintiff can represent. *See Hawkins v. Comparet-Cassani*, 251 F.3d 1230, 1238 (9th Cir. 2001)

13 ("A named plaintiff cannot represent a class alleging … claims that the named plaintiff does not

14 have standing to raise.").

15 The Amended Complaint also includes prayers for punitive damages in connection with

16 their interference with economic advantage, unjust enrichment, unfair competition, and

17 negligence claims, which should independently be dismissed pursuant to Rule 12(b)(6). Under

18 California law, punitive damages are available only upon a showing of "oppression, fraud, or

19 malice," which require some form of intentionally injurious, "despicable," or intentionally

20 misleading conduct. Cal. Civ. Code § 3294(a), (c); *see GBTI, Inc. v. Ins. Co. of Pa.*, No. CV F

21 09-1173, 2009 WL 2365409, *6 (E.D. Cal. July 29, 2009) (collecting authorities on pleading

22 requirements for punitive damages under § 3294). Plaintiffs include no non-conclusory

23 allegations that GitHub or Microsoft engaged in the intentionally wrongful behavior required, so

24 their punitive damages request should be dismissed under Rule 12(b)(6). *E.g.*, *Rhynes v. Stryker*

25 *Corp.*, No. 10-5619 SC, 2011 WL 2149095, *6 (N.D. Cal. May 31, 2011) (dismissing punitive

26 damages request based only on "conclusory allegations"); *Vermillion v. Corrections Corp. of*

27 *Am.*, No. CV F 08-1069, 2008 WL 4058063, *11 (E.D. Cal. Aug. 28, 2008) (similar).

28

II.   PLAINTIFFS' DMCA CLAIMS FAIL BECAUSE PLAINTIFFS DO NOT ALLEGE REMOVAL OR ALTERATION OF CMI FROM IDENTICAL COPIES OF WORKS.

In resolving Defendants' previous motions to dismiss Plaintiffs' copyright management information claims, this Court granted the motions as to § 1202(a) and (b)(2), but denied the motions as to § 1202(b)(1) and (b)(3)—both of which involve removal or alteration of CMI. Order 16, 18-21, ECF No. 95. Plaintiffs have not attempted to amend the dismissed claims, repleading only their removal or alteration claims. Am. Compl. ¶¶ 183-213. GitHub and Microsoft preserve all rights as to those claims, but do not seek to relitigate issues this Court has already resolved. GitHub and Microsoft respectfully ask the Court to address a defect in Plaintiffs' claims that it did not reach in its previous order, and one that is only more pronounced in light of Plaintiffs' Amended Complaint: § 1202(b) claims lie only when CMI is removed or altered from an identical copy of a copyrighted work. *See* Mots. To Dismiss 13-14, ECF No. 50 at 13-14 (Microsoft and GitHub's motions arguing that removal from an "identical copy of the work" is required). Plaintiffs fail to plead this identical copy requirement as to any of their claimed works.

Courts have consistently rejected § 1202(b) claims where the copy from which CMI is allegedly removed is merely an excerpt or modification of the original copy. This includes § 1202(b) claims based on mere "framing" of a photograph in a way that does not include CMI, *id.*; excerpting lecture notes and study questions from textbooks without reproducing CMI, *Faulkner Press, L.L.C. v. Class Notes, L.L.C.*, 756 F. Supp. 2d 1352, 1356, 1359 (N.D. Fla. 2010); copying "aspects" of architectural works but "omitting" the plaintiff's CMI, *Design Basics, LLC v. WK Olson Architects, Inc.*, No. 17 C 7432, 2019 WL 527535, at *5 (N.D. Ill. Feb. 11, 2019); *Frost-Tsuji Architects v. Highway Inn, Inc.*, No. CIV. 13-00496, 2015 WL 263556, at *3 (D. Haw. Jan. 21, 2015), *aff'd*, 700 F. App'x 674 (9th Cir. 2017); or incorporating the underlying content from the original copy into some different form or distinct work without CMI, *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999) (thumbnail versions of images)*, rev'd on other grounds*, 336 F.3d 811 (9th Cir. 2003).

This identical-copy requirement makes sense in light of the overarching purpose of

GITHUB AND MICROSOFT'S MTD AM. COMPL.
NO. 4:22-CV-6823-JST

1   § 1202(b). That statute was enacted to create so-called paracopyright protections, "assist[ing] in

2   tracking and monitoring uses of copyrighted works, as well as licensing of rights and indicating

3   attribution, creation and ownership." S. Rep. No. 105-190 (1998) at 16. That is why CMI is

4   defined not as any information conveyed with any content, but specifically as "information

5   conveyed in connection with copies … of a work." 17 U.S.C. § 1202(c). Were it otherwise, what

6   was intended as a narrow paracopyright protection would swallow copyright protection whole,

7   allowing plaintiffs to plead "removal" of CMI from mere snippets of a work that may have no

8   claim to any copyright protection at all.

9   Plaintiffs' Amended Complaint lays this danger bare. In their initial Complaint, Plaintiffs

10   were coy about what sort of code they were suggesting Copilot might emit without attribution in

11   violation of § 1202(b). But the Amended Complaint's examples show how broad—and legally

12   untenable—their theory is. In the example pertaining to Doe 2, Plaintiffs caused Copilot to emit a

13   list of the ██

14   ████████████. Am. Compl. ¶¶ 100-04. Plaintiffs do not allege that this list of ██████████

15   constitutes a full, identical copy of a work in which they have an interest. Nor could they claim

16   rights in it, since this "code" is merely a table of constants, which constants happen to be the

17   ████████████████████████ and which table happens to mirror the one contained in the Wikipedia

18   article ██

19   ████████████████. See http://en.wikipedia.org/wiki/████████████████████████████. Yet

20   Plaintiffs seem to be suggesting that if Doe 2 is not credited as the author of this obviously

21   uncopyrightable code snippet, Defendants have violated § 1202(b).

22   This fundamental defect is endemic to Plaintiffs' theory of § 1202(b) liability. As the

23   Amended Complaint acknowledges, Copilot's output is limited to short passages of code

24   representing suggested completions for coding processes initiated by a user. *Supra* 4. Copilot's

25   output will therefore virtually never constitute an identical copy of a work. Moreover, the

26   Amended Complaint acknowledges that suggested snippets are "more often" "modified or

27   adapted"—that is, expressed with differences based on the patterns the model has internalized

28   across the training set. Plaintiffs' § 1202(b) theory would embrace *all of this*, ostensibly by

14

1   treating every output Copilot generates that *resembles* code in Plaintiffs' repository as a

2   "Derivative Work" from which CMI has been removed.

3   That theory is wrong on the law. As the case law above demonstrates, § 1202(b) is about

4   identical "copies … of a work"—not about stray snippets and adaptations. And Plaintiffs' theory

5   also cannot be squared with the Amended Complaint's own allegations about how Copilot

6   actually works. Though Plaintiffs' brand Copilot a rote copy-paste machine that cannot

7   understand code, their examples actually show that it does not *copy* anything, but rather generates

8   suggestions afresh based on what it has discerned across a training set. (How else would it decide

9   to use the terms ███

10   ███████████████████████? *See* Am. Compl. ¶ 103 (Plaintiffs acknowledging these

11   variations on Copilot output).) This Court should dismiss the § 1202(b) claims because Plaintiffs

12   have not alleged that Copilot removes CMI from identical copies of Plaintiffs' works.

13   **III.   PLAINTIFFS' TORT CLAIMS ARE PREEMPTED BY THE COPYRIGHT ACT.**

14   Plaintiffs' state-law tort claims are all preempted by the Copyright Act because they are

15   built on the allegation that Defendants, without authorization, copied, distributed, and created

16   derivative works from code in Plaintiffs' repositories. This includes Plaintiffs' claims for tortious

17   and negligent interference with economic advantage (Counts 4 and 5), unjust enrichment and

18   negligence (Counts 6 and 8), and unfair competition to the extent based on state law tort claims

19   (Count 7, Am. Compl. ¶ 276(c), (d)).

20   Section 301 of the Copyright Act preempts "legal or equitable rights that are equivalent to

21   any of the exclusive rights within the general scope of copyright … and come within the subject

22   matter of copyright." 17 U.S.C. § 301(a). Courts evaluate copyright preemption under a "two-

23   part test," holding state law claims preempted where (1) "the 'subject matter' of the state law

24   claim falls within the subject matter of copyright as described in 17 U.S.C. §§ 102 and 103"; and

25   (2) "the rights asserted under state law are equivalent to the rights contained in 17 U.S.C. § 106."

26   *Maloney v. T3Media, Inc.*, 853 F.3d 1004, 1010 (9th Cir. 2017) (quotation marks omitted).

27   The subject matter requirement of the test is plainly satisfied. Plaintiffs' tort claims all

28   involve Plaintiffs' claimed copyright interests in computer code, which constitutes a "literary

1   work" under § 102. *See Google*, 141 S. Ct. at 1196. These claims therefore involve the "subject

2   matter" of copyright, a result that obtains even if the code is not actually protected by the

3   Copyright Act. *See Montz v. Pilgrim Films & Television, Inc.*, 649 F.3d 975, 979 (9th Cir. 2011)

4   (en banc) ("[T]he scope of the subject matter of copyright law is broader than the protections it

5   affords.").

6   The equivalent-rights prong is also met with respect to each claim. When Defendants

7   raised preemption in response to various claims in the initial Complaint, Plaintiffs did not even

8   respond—doubtless because their claims quoted nearly verbatim from the Copyright Act's list of

9   exclusive rights. *See* Order 17, ECF No. 95 ("Plaintiffs' opposition generally does not address

10   Defendants' preemption arguments."). This Court recognized as much in finding Plaintiffs'

11   unjust enrichment claim preempted: "Plaintiffs base their unjust enrichment claim on

12   Defendants' reproduction of Plaintiffs' code as output and Defendants' preparation of derivative

13   works, both of which are rights protected under the Copyright Act." *Id.* at 17-18. Although

14   Plaintiffs replead their tort claims without using copyright language, the substance of the

15   allegations shows that Plaintiffs' tort claims remain mere copyright claims in disguise.

16   Plaintiffs' interference with economic advantage claims are not based on any existing or

17   specific economic opportunity. Plaintiffs claim that GitHub and Microsoft are interfering with

18   their rights to form contracts with the "global open-source community" by "emitting code subject

19   to open-source licenses without the licenses attached." Am. Compl. ¶ 250. This claim at bottom

20   is based on the theory that Plaintiffs have a tort-based exclusive right to control the reproduction

21   and distribution of certain code to the general public, and that Defendants are interfering with

22   those rights. That is a copyright claim. *See Crafty Prods., Inc. v. Fuqing Sanxing Crafts Co.*, 839

23   F. App'x 95, 98-99 (9th Cir. 2020) (tortious interference claim preempted); *Harper & Row*

24   *Publishers, Inc. v. Nation Enters.*, 723 F.2d 195, 201 (2d Cir. 1983) (tortious interference claim

25   preempted where "unauthorized publication is the gravamen of the[] claim"), *rev'd on other*

26   *grounds*, 471 U.S. 539 (1985).

27   It makes no difference that Plaintiffs also append the qualifier "without the licenses

28   attached." Am. Compl. ¶ 250. That is just another way of saying that Plaintiffs have not

GITHUB AND MICROSOFT'S MTD AM. COMPL.
NO. 4:22-CV-6823-JST

authorized the reproduction or distribution of code in the manner they claim Defendants have

engaged in. The claims are still, at bottom, based on reproduction and distribution—they are not

qualitatively different merely because Plaintiffs also allege that the reproduction and distribution

was unauthorized. Were it otherwise, a plaintiff could use state-tort law to dramatically curtail

public domain rights the Copyright Act seeks uniformly to protect. As the Supreme Court has

explained, absent copyright protection, "the public may use the … work at will and without

attribution." *Dastar Corp. v. Twentieth Century Fox Film Corp.*, 539 U.S. 23, 34 (2003).

Plaintiffs' rights to control reproduction and distribution to the "global … community" therefore

lie in the Copyright Act or perhaps contract—but they cannot be derived from state tort-law.

Plaintiffs' unjust enrichment and negligence claims, meanwhile, are based on the

exclusive right "to prepare derivative works." 17 U.S.C. § 106. Both claims appear to target the

training of Codex and Copilot. Am. Compl. ¶ 269 ("Defendants have unjustly utilized access to

Licensed Materials hosted on GitHub to create Codex and Copilot."); *id.* ¶ 271 ("Plaintiffs did not

consent to the unauthorized use of their Licensed Materials to train Codex and Copilot."); *id.*

¶ 284 ("Defendants breached their duties by … engineering, designing, maintaining, and

controlling systems—including Codex and Copilot—which are trained on Plaintiffs' and Class

members' Licensed materials without their authorization."). Plaintiffs explicitly allege,

moreover, that Copilot is a derivative work: "The definition [of Derivative Works] also includes

the Copilot product itself, which is a Derivative Work based upon a large corpus of Licensed

Materials." Am. Compl. ¶ 194 n.34. So, by the terms of Plaintiffs' own allegations, the conduct

underlying their unjust enrichment and negligence claims is the preparation of a derivative work.

It is therefore preempted. *See* Order 17, ECF No. 95 (finding preemption of unjust enrichment

claim based upon preparation of a derivative work).

Lastly, insofar as Plaintiffs' UCL claim is predicated on the above state-law torts, it too is

preempted. *See Maloney*, 853 F.3d at 1019 (holding UCL claim preempted); *Kodadek v. MTV

Networks, Inc.*, 152 F.3d 1209, 1213 (9th Cir. 1998) (same).

IV.   THE TORT CLAIMS ARE DEFECTIVE FOR OTHER REASONS AS WELL.

A.   Plaintiffs Fail To Identify Applicable State Law For Their Common Law Claims.

In its order on Defendants' motions to dismiss the initial Complaint, this Court noted Plaintiffs' failure to "identify the state law which applies to each" of its common law tort claims and directed Plaintiffs "in any future amended complaint … [to] identify the state under whose law the claim is brought." Order 4 n.5, ECF No. 95. Plaintiffs have failed to comply with this Court's directive with respect to their common law tort claims for intentional interference with prospective economic relations, negligent interference with prospective economic relations, unjust enrichment, unfair competition, and negligence. This is "grounds for dismissal." *Id.* (quoting *In re Nexus 6P Prods. Liab. Litig.*, 293 F. Supp. 3d 888, 933 (N.D. Cal. 2018) (quotation marks omitted)).

B.   Plaintiffs Fail To State A Claim For Intentional Or Negligent Interference With Economic Relations.

Plaintiffs do not replead their original tortious interference with contractual relations claim. Instead, they replace it with claims for intentional and negligent interference with prospective economic relations (Counts 4 and 5). Although "[t]ortious interference with prospective economic advantage … does not depend on the existence of a legally binding contract," it still requires a plaintiff to show that the defendant interfered with a specific "economic relationship between the plaintiff and some third party, which carries the probability of future economic benefit to the plaintiff." *Ixchel Pharma, LLC v. Biogen, Inc.*, 9 Cal. 5th 1130, 1141 (2020) (alterations adopted). For both claims, Plaintiffs must identify specific, known relationships and plausibly allege the probability of concrete future economic benefit from the identified relationship. *Westside Ctr. Assocs. v. Safeway Stores 23, Inc.*, 42 Cal. App. 4th 507, 523-28 (1996); *Rosen v. Uber Techs., Inc.*, 164 F. Supp. 3d 1165, 1178-79 (N.D. Cal. 2016). The only "difference between intentional interference and negligent interference with prospective economic advantage relates to the defendant's intent." *Crown Imports, LLC v. Super. Ct.*, 223 Cal. App. 4th 1395, 1404 n.10 (2014).

Here, the Amended Complaint does not identify any specific business expectancy or third-

1   party relationship with which GitHub and Microsoft allegedly interfered. The Amended

2   Complaint merely alleges that Defendants affected Plaintiffs' relationships with the "global open-

3   source community" and interfered with Plaintiffs' vague expectation that unidentified

4   "programmers would use, modify, copy or otherwise iterate on their posted code subject to the

5   terms of the open-source licenses the code was published subject to." Am. Compl. ¶¶ 247, 258,

6   260. It is well-settled, however, that a plaintiff cannot base an economic interference claim on

7   abstract interference with opportunity in "the market" generally or on vague "lost

8   opportunit[ies]." *Westside Ctr. Assocs.*, 42 Cal. App. 4th at 527. Instead, Plaintiffs must identify

9   specific relationships with specific third-parties and plausibly allege that defendant "knew of the

10   existence of the relationship." *Id.* at 526; *see O'Connor v. Uber Techs., Inc.*, 58 F. Supp. 3d 989,

11   998 (N.D. Cal. 2014) ("[I]nterference with potential" third parties "with whom the plaintiff did

12   not have an existing relationship generally is not sufficient to state a claim.") Allowing the claim

13   to proceed based on Plaintiffs' allegation that "programmers" generally would use open-source

14   code differently absent Defendants' actions would "allow[] recovery no matter how speculative

15   the plaintiff's expectancy." *Rosen*, 164 F. Supp. 3d at 1179.

16   Nor do Plaintiffs plausibly allege that these unspecified relations in fact carried the

17   likelihood of generating "future economic benefit." *Ixchel*, 9 Cal. 5th at 1141. Plaintiffs vaguely

18   allege that Defendants caused them to miss out on "substantial benefits" that come with "the

19   creation and distribution of open-source code subject to these open-source licenses," Am. Compl.

20   ¶ 243; "prevented Copilot users from becoming part of the user communities that would

21   ordinarily accrete around the open-source projects of Plaintiffs and the Class," Am. Compl.

22   ¶ 249; and prevented them from "optimiz[ing] the likelihood of accruing communities of other

23   GitHub customers for their own projects," Am. Compl. ¶ 256. But the Amended Complaint does

24   not allege how actual "future economic benefits" would have flown to Plaintiffs from their

25   "becoming part of the user communities" or engaging in "open-source distribution" in the

26   absence of Defendants' challenged conduct. Instead, it "assumes what normally must be proved,

27   i.e., that it is reasonably probable the plaintiff would have received" a concrete "expected benefit

28   had it not been for the defendant's interference." *Rosen*, 164 F. Supp. 3d at 1179. The Amended

GITHUB AND MICROSOFT'S MTD AM. COMPL.
                                                          NO. 4:22-CV-6823-JST

1   Complaint's assertions that Defendants "deprive[d] Plaintiffs of the economic benefits of open-

2   source distribution," Am. Compl. ¶ 262, and "the future economic benefits likely to arise from

3   those relationships," Am. Compl. ¶ 246, are thus paradigmatic conclusory allegations that fail to

4   state a claim. *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009).

5   The Amended Complaint also fails to plausibly allege "actual disruption" or specific

6   "harm," two additional elements of the claims. *Korea Supply Co. v. Lockheed Martin Corp.*, 29

7   Cal. 4th 1134, 1153 (2003). As a result, the Amended Complaint does not state a claim for any

8   economic interference tort.

9   **C.   Plaintiffs' New Attempt At An Unjust Enrichment Claim Does Not Cure Its**
              Defects.
10

11   Under California law, the only form of independent claim sounding in unjust enrichment

12   is "a quasi-contract claim seeking restitution." *ESG Cap. Partners, LP v. Stratos*, 828 F.3d 1023,

13   1038 (9th Cir. 2016) (citation omitted). "However, 'to allege a quasi-contract claim, a party must

14   plausibly allege the absence of any applicable and enforceable contract provisions, even if in the

15   alternative.'" *In re Bang Energy Drink Mktg. Litig.*, No. 18-cv-05758, 2020 WL 4458916, at *10

16   (N.D. Cal. Feb. 6, 2020) (quoting *City of L.A. v. Sprint Sols., Inc.*, No. 17-cv-00811, 2019 WL

17   5568879, at *9 (E.D. Cal. Oct. 29, 2019)). But Plaintiffs have not anywhere "alternatively

18   alleged the absence of such provisions," *id.*, so this claim should be dismissed for that reason

19   alone.

20   The claim additionally fails for the reason that Plaintiffs have not alleged how Defendants

21   benefited at their expense. *See Stratos*, 828 F.3d at 1038 ("To allege unjust enrichment as an

22   independent cause of action, a plaintiff must show that the defendant received and unjustly

23   retained a benefit at the plaintiff's expense."). Plaintiffs say that "[b]y using Plaintiffs' Licensed

24   Materials to train Codex and Copilot, Plaintiffs and the Class were deprived of the benefits of

25   their open-source licenses, including monetary damages." Am. Compl. ¶ 270. But Plaintiffs

26   have not alleged any injury based on training. *See also* Order at 8 n.7, ECF No. 95 ("Plaintiffs

27   do not allege they were injured by Defendants' use of licensed code as training data[.]"). Even

28   construed as a claim based on output, the unjust enrichment claim fails because Plaintiffs have

1   not alleged that any Copilot user other than themselves has ever prompted Copilot to suggest code

2   that is a copy of theirs. *See supra* 9-10.

3   **D.**   **<u>Plaintiffs Have Not Alleged The Deprivation Of Money Or Property Required</u>**
                      <u>To Make A UCL Claim</u>.

4

5   Plaintiffs duplicate their DMCA, open-source license, and tort claims as purported

6   violations by GitHub of California's Unfair Competition Law, Cal. Bus. & Prof. Code §§ 17200,

7   *et seq.* (Count 7). This is dead on arrival. The "UCL's standing requirements" are "more

8   stringent than the federal standing requirements" and require a showing of "lost money or

9   property." *Kwikset Corp. v. Super. Ct.*, 51 Cal. 4th 310, 324 (2011); Cal. Bus. & Prof. Code

10   § 17204.

11   The Amended Complaint does not plausibly allege the required "lost money or property."

12   Cal. Bus. & Prof. Code § 17204. It asserts that "Plaintiffs and the Class have suffered economic

13   injury as a result of Defendants' conduct." Am. Compl. ¶ 281. But the Amended Complaint does

14   not allege any factual theory of *how* Plaintiffs suffered monetary loss as a result of the training of

15   Copilot or its suggestions, let alone any specific allegations supporting such a theory. All they

16   state is that "there are economic benefits to the creation of open-source works such as generating

17   market share for programs, increasing national or international reputation by incubating open-

18   source projects, and deriving value from improvements to software based on suggestions by end-

19   users." Am. Compl. ¶ 281. This general observation provides no explanation of how

20   Defendants' conduct caused them to lose "money or property." Plaintiffs therefore offer only a

21   "[t]hreadbare recital[]" of the UCL's "stringent" standing element "devoid of 'further factual

22   enhancement,'" which "do[es] not suffice." *Iqbal*, 556 U.S. at 678 (citation omitted).

23   **E.**   **<u>Plaintiffs' Negligence Claim Fails To Identify Any Legal Duty Sounding In</u>**
                      <u>Tort</u>.

24

25   Plaintiffs' negligence claim should be dismissed for lack of plausible allegations of duty

26   and injury. "The existence of a duty of care owed by a defendant to a plaintiff is a prerequisite to

27   establishing a claim for negligence." *Langan v. United Servs. Auto. Ass'n*, 69 F. Supp. 3d 965,

28   987 (N.D. Cal. 2014) (quoting *Nymark v. Heart Fed. Savings & Loan Ass'n.,* 231 Cal.App.3d

                          GITHUB AND MICROSOFT'S MTD AM. COMPL.
                          NO. 4:22-CV-6823-JST

1   1089, 1095 (1991)). Plaintiffs allege "Microsoft and GitHub also owed its user a duty of care not

2   to itself use the Licensed Materials in a way that would foreseeably cause Plaintiffs and Class

3   members injury, for instance, by using Licensed Materials to train Copilot." Am. Compl. ¶ 286.

4   This apparent duty is grounded in "Defendants' relationship" to Plaintiffs, including "Defendants'

5   contractual obligations, custom and practice, right to control information in its possession,

6   exercise of control over the information in its possession, authority to control the information in

7   its possession, and the commission of affirmative acts that resulted in said harms and losses."

8   Am. Compl. ¶ 283. But *all of* these ostensible sources of duty are in fact grounded in the

9   "contractual obligations" (and rights) that define the terms of users' relationship with GitHub.

10   And that forecloses any tort-based duty, because it is settled California law that a duty in tort

11   cannot be derived from obligations in a contract absent some "special relationship" recognized in

12   law—which Plaintiffs do not and cannot claim here. *See Foley v. Interactive Data Corp.*, 47 Cal.

13   3d 654, 682-94 (1988).

14   Moreover, any alleged violation of duty predicated on *training* Copilot with code

15   published in public GitHub repositories, it is expressly *foreclosed* by GitHub's Terms of Service.

16   Am. Compl. Ex. 1. GitHub's TOS expressly authorizes the training of Copilot. The TOS

17   informs users that "You own content you create, but you allow us certain rights to it." Am.

18   Compl. Ex. 1 at 27 (TOS at 6). "These license grants apply to Your Content" when uploaded to

19   GitHub, notwithstanding any other license terms that might be attached. *Id.* Every GitHub user

20   chooses whether to allow their "repositories to be viewed publicly." *Id.* No user is required to

21   make a repository public. If they choose to do so, they "grant [GitHub] … the right to store,

22   archive, parse, and display" publicly posted content and "make incidental copies, as necessary to

23   provide the Service, including improving the Service over time." *Id.* This includes the right to

24   "copy" public material "to our database," "parse it into a search index or otherwise analyze it on

25   our servers," and "share it with other users." *Id.* at 27-28 (TOS at 6-7). The "Service" includes

26   all "the applications, software, products, and services provided by GitHub," *id.* at 24 (TOS at 3),

27   which includes Copilot. *See* Am. Compl. ¶ 8; Am. Compl. Ex. 1 at 23, 37 (TOS at 3, GitHub

28   Copilot Terms).

As Plaintiffs allege, Copilot functions by evaluating the code contained in GitHub's public repositories, Am. Compl. ¶¶ 86-87, in order to "infer[] statistical patterns governing the structure of code." Am. Compl. ¶ 56. That activity is squarely encompassed by the contractual authorization, from every public code repository owner, for GitHub to "store, archive, parse, and display" publicly posted content "as necessary to provide … the applications, software, products, and services provided by GitHub" and to "analyze it on our servers." Am. Compl. Ex. 1 at 26-27, 23 (TOS at 6-7, 3). Even if the Amended Complaint put forward any plausible theory of how inspection and analysis of publicly posted material violated an open source license provision, Plaintiffs' acceptance of the GitHub TOS forecloses a claim for breach of contract based on Copilot's training on the public code. "Dismissal is … proper where the complaint alleges facts that demonstrate that the [claim] is barred as a matter of law." *Orchard Supply Hardware LLC v. Home Depot USA, Inc.*, 967 F. Supp. 2d 1347, 1354 (N.D. Cal. 2013). So it is here.

Plaintiffs also fail to allege any injury. Plaintiffs assert Defendants' "breach of their duties," "by using open-source code in violation of open-source licenses to train Codex and Copilot," "proximately caused Plaintiffs and Class members injuries," Am. Compl. ¶¶ 287-88, but they do "not allege *how*" they were "injured by such conduct nor what damages [Plaintiffs] sustained from such conduct." *West v. City & Cnty. of S.F.*, No. 21-cv-02370, 2022 WL 1556415, at *13 (N.D. Cal. May 17, 2022) (emphasis added). Count 8 should be dismissed.

CONCLUSION

The Court should grant the motions to dismiss portions of the Amended Complaint.

Dated: June 29, 2023

Orrick, Herrington & Sutcliffe LLP

By: _____/s/ Annette L. Hurst_____
ANNETTE L. HURST
Attorneys for Defendants
GitHub, Inc. and Microsoft Corp.

GITHUB AND MICROSOFT'S MTD AM. COMPL.
NO. 4:22-CV-6823-JST

EXHIBIT H

1   MICHAEL A. JACOBS (SBN 111664)
    MJacobs@mofo.com
2   JOSEPH C. GRATZ (SBN 240676)
    JGratz@mofo.com
3   TIFFANY CHEUNG (SBN 211497)
    TCheung@mofo.com
4   MELODY E. WONG (SBN 341494)
    MelodyWong@mofo.com
5   MORRISON & FOERSTER LLP
    425 Market Street
6   San Francisco, California 94105-2482
    Telephone: (415) 268-7000
7   Facsimile: (415) 268-7522
    [CAPTION PAGE CONTINUED ON NEXT PAGE]
8
    Attorneys for Defendants OPENAI, INC., a Delaware nonprofit
9   corporation, OPENAI, L.P., a Delaware limited partnership,
    OPENAI OPCO, L.L.C., a Delaware limited liability company,
10 OPENAI GP, L.L.C., a Delaware limited liability company,
    OPENAI STARTUP FUND GP I, L.L.C., a Delaware limited
11 liability company, OPENAI STARTUP FUND I, L.P., a
    Delaware limited partnership, OPENAI STARTUP FUND
12 MANAGEMENT, LLC, a Delaware limited liability company

13

UNITED STATES DISTRICT COURT

14

NORTHERN DISTRICT OF CALIFORNIA

15

16 | J. DOE 1 and J. DOE 2, individually and on behalf of all others similarly situated, | Case No. 4:22-cv-06823-JST
     Case No. 4:22-cv-07074-JST

17

| Plaintiffs, | Hon. Jon S. Tigar

18

19 | v. | **CLASS ACTION**

20 | GITHUB, INC., a Delaware corporation; MICROSOFT CORPORATION, a Washington corporation; OPENAI, INC., a Delaware nonprofit corporation; OPENAI, L.P., a Delaware limited partnership; OPENAI OPCO, L.L.C., a Delaware limited liability company; OPENAI GP, L.L.C., a Delaware limited liability company; OPENAI STARTUP FUND GP I, L.L.C., a Delaware limited liability company; OPENAI STARTUP FUND I, L.P., a Delaware limited partnership; OPENAI STARTUP FUND MANAGEMENT, LLC, a Delaware limited liability company, | **DEFENDANTS OPENAI, INC., OPENAI, L.P., OPENAI OPCO, L.L.C., OPENAI GP, L.L.C., OPENAI STARTUP FUND GP I, L.L.C., OPENAI STARTUP FUND I, L.P., AND OPENAI STARTUP FUND MANAGEMENT, LLC'S MOTION TO DISMISS FIRST AMENDED COMPLAINT**

21

22

23

24

25

26 | Defendants. | **[PUBLIC VERSION OF DOCUMENT SOUGHT TO BE FILED UNDER SEAL]**

Date: September 14, 2023
Time: 2:00 p.m.
Dept: Courtroom 6

27

28

1   ALLYSON R. BENNETT (SBN 302090)
    ABennett@mofo.com
2   ROSE S. LEE (SBN 294658)
    RoseLee@mofo.com
3   ALEXANDRA M. WARD (SBN 318042)
    AlexandraWard@mofo.com
4   MORRISON & FOERSTER LLP
    707 Wilshire Boulevard
5   Los Angeles, California 90017-3543
    Telephone: (213) 892-5200
6   Facsimile: (213) 892-5454

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

1

NOTICE OF MOTION AND MOTION

2

TO ALL PARTIES AND THEIR ATTORNEYS OF RECORD:

3

PLEASE TAKE NOTICE that on September 14, 2023 at 2:00 p.m., or at a different time

4

and date set by the Court, Defendants OPENAI, INC., OPENAI, L.P., OPENAI OPCO, L.L.C.,

5

OPENAI GP, L.L.C., OPENAI STARTUP FUND GP I, L.L.C., OPENAI STARTUP FUND I,

6

L.P. AND OPENAI STARTUP FUND MANAGEMENT, LLC (hereinafter "OPENAI"), by and

7

through counsel, will and hereby do move the Court to dismiss the following claims asserted in

8

the First Amended Complaint against OpenAI: (1) violation of the Digital Millennium Copyright

9

Act ("DMCA"), 17 U.S.C. § 1202, *et seq.*; (2) intentional interference with prospective economic

10

relations; (3) negligent interference with prospective economic relations; (4) unjust enrichment

11

under common law; (5) unfair competition under common law and California Business &

12

Professions Code § 17200, *et seq*.; and (6) negligence.

13

This Motion is made pursuant to Federal Rules of Civil Procedure 12(b)(1) and 12(b)(6),

14

and is based upon this Notice of Motion and Motion, the Memorandum of Points and Authorities

15

included herewith, the Proposed Order submitted herewith, all pleadings and papers on file in this

16

action, and such further evidence that may be submitted to the Court or before the hearing.

17

18

19

20

21

22

23

24

25

26

27

28

TABLE OF CONTENTS

1

<u>TABLE OF AUTHORITIES</u>

2

Page(s)

3

Cases

28

Constitution, Statutes, and Rules

STATEMENT OF THE ISSUES TO BE DECIDED

This motion raises the following issues:

1.   **Article III Standing.** Whether Plaintiffs' claims for damages should be dismissed for lack of standing under Article III of the United States Constitution.

2.   **Preemption.** Whether the Copyright Act preempts the state law causes of action for intentional interference with prospective economic relations, negligent interference with prospective economic relations, unjust enrichment, unfair competition, and negligence.

3.   **Digital Millennium Copyright Act ("DMCA"), 17 U.S.C. §§ 1201-1205 (Count I).** Whether this claim should be dismissed for failing to allege (i) removal of CMI from identical copies or (ii) distribution of works with removed CMI.

4.   **Intentional Interference with Prospective Economic Relations (Count IV).** Whether this claim should be dismissed for failure to plead the existence of an economic relationship between Plaintiffs and a third party with a probability of economic benefit, OpenAI's knowledge of the existence of that relationship, intentional acts on the part of OpenAI designed to disrupt the relationship, actual disruption of the relationship, and economic harm to Plaintiffs as a result.

5.   **Negligent Interference with Prospective Economic Relations (Count V).** Whether this claim should be dismissed for failure to plead that OpenAI owed Plaintiffs a duty of care, as well as other required elements.

6.   **Unjust Enrichment (Count VI).** Whether this claim should be dismissed because there is no standalone cause of action for unjust enrichment and Plaintiffs fail to plead the required elements.

7.   **Unfair Competition (Count VII)**. Whether this claim should be dismissed for failure to (i) sufficiently plead a lack of an adequate legal remedy or an economic injury for the Unfair Competition Law ("UCL") claim under California Business. & Professions Code § 17200; (ii) plead a predicate violation for the "unlawful" claim; (iii) allege sufficient facts in support of the "unfair" claim; (iv) satisfy the heightened pleading requirements for fraud and allege reliance for the "fraudulent" claim; and (v) identify an actionable basis for a common law claim.

8.   **Negligence (Count VIII).** Whether this claim should be dismissed for failure to plead that OpenAI owed Plaintiffs a duty of care.

1

MEMORANDUM OF POINTS AND AUTHORITIES

2

I.   INTRODUCTION

3

This Court previously dismissed all of Plaintiffs' claims for monetary damages because

4

Plaintiffs had not plausibly alleged that any Defendant had reproduced protectible aspects of

5

Plaintiffs' code. The Court also dismissed Plaintiffs' state law tort claims as preempted by the

6

Copyright Act. Plaintiffs have amended their complaint, but they have not cured the defects that

7

the Court previously identified with respect to the claims they have re-pled. Their new claims

8

fare no better.

9

First, Plaintiffs still have not plausibly alleged that any of their code was output by Codex

10

or Copilot prior to the filing of the original complaint by anyone other than Plaintiffs themselves.

11

They therefore continue to lack standing to bring claims for monetary relief.

12

Second, Plaintiffs' newly pled facts showing the contortions required to cause Copilot to

13

generate approximations of Plaintiffs' code reinforce why Plaintiffs' DMCA claims fail. The

14

new allegations demonstrate that they cannot plausibly allege removal of CMI from identical

15

copies of Plaintiffs' code.

16

Third, Plaintiffs' state law tort and unjust enrichment claims fail because they continue to

17

be preempted. The Court previously found Plaintiffs' unjust enrichment claim preempted by the

18

Copyright Act because Plaintiffs alleged only the unauthorized reproduction of Plaintiffs' code.

19

That's still true of the first amended complaint—for both the unjust enrichment and the tort

20

claims. Plaintiffs' unjust enrichment and tort claims also suffer from a host of legal deficiencies

21

on the merits.

22

Fourth, Plaintiffs' unfair competition claim fails because (among other reasons) Plaintiffs

23

still have not pled an inadequate legal remedy, or a predicate violation that caused them an

24

economic injury.

25

Accordingly, Open AI's motion to dismiss should be granted.

26

II.   FACTUAL AND PROCEDURAL BACKGROUND

27

The Original Complaint. This case was originally brought by four named Plaintiffs

28

(proceeding anonymously). They alleged that materials that they had posted to GitHub under

1   open-source licenses were used as part of the training set for Codex and Copilot, although they

2   did not allege that any of their copyrighted code had been output by Codex or Copilot. (*See, e.g.,*

3   Compl. (Dkt. No. 1) ¶¶ 90-91.) As a result, they brought claims against Microsoft Corporation

4   ("Microsoft"), GitHub, Inc. ("GitHub"), and six OpenAI entities (together with subsequently

5   named defendant OpenAI OpCo, L.L.C., "OpenAI," and, collectively with Microsoft and GitHub,

6   "Defendants").

7   **The Court's Prior Order.** On January 26, 2023, Defendants moved to dismiss the

8   complaint, both on standing grounds and on the merits. (Defs.' Mots. to Dismiss (Dkt. Nos. 50,

9   53).) On May 11, 2023, this Court granted in part and denied in part Defendants' motions to

10   dismiss. (Order Granting in Part and Denying in Part Defs.' Mot. to Dismiss (Dkt. No. 95,

11   "Order").) The Court denied Defendants' motion with respect to Plaintiffs' claims for (1) breach

12   of contract and (2) declaratory relief under Sections 1201(b)(1) and 1201(b)(3) of the DMCA.

13   But it granted Defendants' motions on all other claims.

14   With respect to Plaintiffs' claims for monetary relief, the Court found that Plaintiffs

15   lacked standing because "Plaintiffs [did] not identify any instance of Copilot reproducing

16   Plaintiffs' licensed code and therefore failed to plead a particularized injury sufficient to confer

17   standing." (*Id.* at 7.) The Court also found that Plaintiffs lacked standing for their privacy-based

18   claims (violations of the CCPA and negligence) and dismissed them. (*Id.*)

19   On the merits:

20   1.   Plaintiffs' claims under Section 1202(a) of the DMCA, tortious interference with a

21   contract, false designation of origin, and unfair competition claims under the Lanham Act and

22   common law were dismissed because Plaintiffs had failed to defend them in their opposition. (*Id.*

23   at 16.)

24   2.   Plaintiffs' unjust enrichment claim was dismissed as preempted by the Copyright

25   Act. (*Id.* at 16-18.)

26   3.   Plaintiffs' UCL claim was dismissed because it was predicated largely on other

27   claims that had been dismissed. (*Id.* at 22-23.) The only plausibly surviving predicate claims

28   were violations of Sections 1202(b)(1) and (b)(3), but for those claims, the Court held that

1   Plaintiff failed to show "how Defendants' alleged violations of the DMCA have caused or will

2   cause Plaintiffs economic injury." (*Id.* at 23.)

3   4.   Plaintiffs' civil conspiracy and declaratory relief claims were dismissed because

4   neither is a standalone cause of action. (*Id.* at 24.)

5   5.   Plaintiffs' DMCA claim under Section 1202(b)(2) was dismissed based on

6   Plaintiffs' failure to allege distribution of altered copyright management information ("CMI").

7   (*Id.* at 21.)

8   **The Amended Complaint.** On June 8, 2023, Plaintiffs filed the FAC, which added both

9   a new OpenAI defendant and a new plaintiff. (First Am. Compl. (Dkt. No. 97-3, "FAC").) The

10 FAC asserts seven causes of action against Open AI, some of which are new and some of which

11 have been re-pled. Plaintiffs again assert claims for (1) violation of Sections 1202(b)(1) and

12 (b)(3) of the DMCA, 17 U.S.C. § 1202, *et seq.* (Count I); (2) breach of contract (Count II); (3)

13 unjust enrichment (Count VI); (4) unfair competition under common law and California Business

14 & Professions Code § 17200, *et seq.* (Count VII); and (5) negligence (Count VIII).[1] (FAC ¶¶

15 183-229, 266-289.)   Plaintiffs added new claims for (1) intentional interference with prospective

16 economic relations (Count III); and (2) negligent interference with prospective economic relations

17 (Count IV). Plaintiffs did not re-plead their claims for violations of Section 1202(a) and

18 1202(b)(2) of the DMCA, tortious interference with a contract, false designation of origin,

19 violation of the CCPA, civil conspiracy and declaratory relief.

20   Plaintiffs' new factual allegations principally address the Court's finding that the original

21 complaint did not plausibly allege reproduction of Plaintiffs' code. (Order at 8.) For Does 1, 2,

22 and 5, the FAC now alleges that Copilot output code they allegedly published to GitHub,

23 although they tellingly omit *when* that output supposedly occurred. Further, Plaintiffs'

24 allegations demonstrate that Plaintiffs themselves went to grant lengths to cause the output that

25 they now complain of.

26 _____

[1] Where the complaint refers to common law or state law, OpenAI applies California law for
27 purposes of this motion. OpenAI does not concede that California law can be applied to acts
    occurring outside California.
28

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

1 ██

2 ██

3 ██

4 ██████████████████████████████████████

5   ██

6 ██

7 ██

8 ██

9 ██

10 ██

11 ██

12 ██

13 ██

14 ██

15 ██

16 ██████

17   Nowhere do Plaintiffs allege that any of these prompts were provided before the original

18 complaint was filed; nor do they allege that anyone other than Plaintiffs input (or would input)

19 those particular prompts.

20 **III.   LEGAL STANDARD**

21   **A.   Motion to Dismiss for Lack of Subject Matter Jurisdiction.**

22   Federal courts "possess only that power authorized by Constitution and statute."

23 *Kokkonen v. Guardian life Ins. Co. of Am.*, 511 U.S. 375, 377 (1994). Article III "limits the

24 jurisdiction of federal courts to 'Cases' and 'Controversies.'" *Lance v. Coffman,* 549 U.S. 437,

25 439 (2007). If a plaintiff lacks Article III standing to bring a suit, the federal court lacks subject

26 matter jurisdiction, and the suit must be dismissed under Rule 12(b)(1). *Cetacean Cmty. v. Bush*,

27 386 F.3d 1169, 1174 (9th Cir. 2004). Once a defendant has moved to dismiss under Rule

28 12(b)(1), the plaintiff bears the burden of establishing the court's jurisdiction. *See Chandler v.*

1   *State Farm Mut. Auto Ins. Co.*, 598 F.3d 1115, 1122 (9th Cir. 2010).

2   **B.   Motion to Dismiss for Failure to State a Claim Under Rule 12(b)(6).**

3   To survive a Rule 12(b)(6) motion to dismiss, "a complaint must contain sufficient factual

4   matter, accepted as true, to state a claim to relief that is plausible on its face." *Ashcroft v. Iqbal*,

5   556 U.S. 662, 678 (2009) (cleaned up). Dismissal is appropriate "where the complaint lacks a

6   cognizable legal theory or sufficient facts to support a cognizable legal theory." *Mendiondo v.*

7   *Centinela Hosp. Med. Ctr.*, 521 F.3d 1097, 1104 (9th Cir. 2008). Moreover, "[f]actual allegations

8   must be enough to raise a right to relief above the speculative level" and "a formulaic recitation of

9   the elements of a cause of action will not do." *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 555

10   (2007). The Court need not accept as true conclusory allegations or legal characterizations, nor

11   need it accept unreasonable inferences or unwarranted factual deductions. *In re Gilead Scis. Sec.*

12   *Litig.*, 536 F.3d 1049, 1055 (9th Cir. 2008).

13   **IV.   ARGUMENT**

14   **A.   Plaintiffs Lack Article III Standing to Assert Their Claims for Monetary**
         Damages.
15
     Plaintiffs' claims for monetary damages must be dismissed because they have again failed
16
     to sufficiently plead that they suffered a cognizable injury to satisfy "the irreducible constitutional
17
     minimum of standing" under Article III. *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560
18
     (1992). As this Court previously held, "[a]n increased risk of future harm alone is not sufficiently
19
     concrete to confer standing for damages." (Order at 8-9 (citing *TransUnion LLC v. Ramirez*, 141
20
     S. Ct. 2190, 2210-11 (2021)).)   The "mere risk of future harm…cannot qualify as a concrete
21
     harm," unless "the exposure to the risk of future harm itself causes a *separate* concrete harm. (*Id.*
22
     (emphasis in original).) And where plaintiffs "do not allege that they themselves have suffered
23
     the injury they describe, they do not have standing to seek retrospective relief for that injury," as
24
     there would be "no injury redressable by the monetary damages they seek." (*Id. at* 8 & n.8.)
25
     Because Plaintiffs did not "allege that they themselves have suffered the injury" nor had
26
     they "alleged any additional, concrete harm associated with [an] increased risk of
27
     misappropriation" in their original complaint, this Court concluded that those allegations "cannot
28

1   provide standing for Plaintiffs' damages claims." (*Id.* at 8-9.) The FAC attempts to cure that

2   deficiency by offering four "specific examples" where Copilot purportedly output code posted to

3   GitHub by Does 1, 2, and 5. (FAC ¶¶ 98-126.) Does 3 and 4 have yet again failed to plead

4   specific instances in which *their* code was output by Copilot. Because Does 3 and 4 fail to raise

5   any new allegations, as this Court previously determined, their claims for monetary damages must

6   be dismissed. (*See* Order at 8-9.) *See also Alsheikh v. Lew*, No. 3:15-CV-03601-JST, 2016 WL

7   1394338, at *2-3 (N.D. Cal. Apr. 7, 2016) (dismissing claim for lack of Article III standing where

8   plaintiff did not "identif[y] any particular injury that he has suffered").

9   Does 1 and 2 also fail to establish standing to seek monetary damages because they have

10   not shown their standing existed at the start of this action. Despite new allegations asserting that

11   Copilot reproduces their code as output when given highly particularized prompts ███████

12   ██

13   ████████████████████████, these plaintiffs have not alleged that Copilot generated this

14   output before this action was commenced. Courts have recognized that "standing is determined

15   as of the *commencement* of the litigation." *Yamada v. Snipes*, 786 F.3d 1182, 1203 (9th Cir.

16   2015) (emphasis added; cleaned up); *see also Lujan*, 504 U.S. at 569 n.4 ("The existence of

17   federal jurisdiction ordinarily depends on the facts *as they exist* when the complaint is filed.")

18   (emphasis added). The "party invoking the jurisdiction of the court cannot rely on events that

19   unfolded after the filing of the complaint to establish its standing." *Wilbur v. Locke*, 423 F.3d

20   1101, 1107 (9th Cir. 2005) (cleaned up). Their failure to allege that their code was output prior to

21   the filing of the original complaint is therefore fatal to their claims.

22   Nor have Does 1, 2, and 5 alleged that Copilot generated this output for anyone other than

23   Plaintiffs. Plaintiffs cannot manufacture their own injuries to satisfy standing by themselves

24   prompting Copilot to reproduce their code. *See Callahan v. Ancestry.com Inc.*, No. 20-CV-

25   08437-LB, 2021 WL 2433893, at *5 (N.D. Cal. June 15, 2021) (finding that plaintiff's research

26   into defendant's use of records to find support for his privacy claims "is not [a] concrete injury"

27   because "no case establishes that investigation untethered to harm reduction is injury in fact.").

28   Accordingly, Plaintiffs lack standing to pursue any claims for monetary damages.

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

B.   The Copyright Act Preempts Several State Law Causes of Action.

Federal law preempts Plaintiffs' claims for intentional and negligent interference with prospective economic relations, unjust enrichment, unfair competition, and negligence, and accordingly, provides a basis for dismissal. Preemption under Section 301 of the Copyright Act applies if (1) "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" and (2) "whether the rights asserted under state law are equivalent to the rights contained in 17 U.S.C. § 106, which articulates the exclusive rights of copyright holders." *Maloney v. T3Media, Inc.*, 853 F.3d 1004, 1010 (9th Cir. 2017) (cleaned up). This Court previously found that Plaintiffs' unjust enrichment claim was preempted because it was based on "Defendants' reproduction of Plaintiffs' code as output and Defendants' preparation of derivative works, both of which are rights protected under the Copyright Act." (Order at 17-18.) For similar reasons, the following state law claims alleged in the first amended complaint are preempted:

- **Intentional and negligent interference with prospective economic relations.** Plaintiffs claim that OpenAI has "deprived" and "disrupt[ed]" "the economic benefits of open-source licenses" by reproducing code without the licenses attached. (FAC ¶¶ 249-52, 257, 261.) These claims, in essence, boil down to an allegation that OpenAI wrongfully copied Plaintiffs' code. A claim that is "predicated on [Defendants'] unauthorized copying of [Plaintiffs'] code" is "not qualitatively different from [a] copyright infringement" claim, and is preempted. *See Media.net Advert. FZ-LLC v. NetSeer, Inc.*, 156 F. Supp. 3d 1052, 1073-74 (N.D. Cal. 2016) (finding preemption of plaintiff's intentional interference with prospective economic advantage where plaintiff alleged that defendant directly copied plaintiff's code to create its own product and "interfered with [p]laintiff's existing contractual relations with Microsoft"); *see also Aghmane v. Bank of Am., N.A.*, No. C-13-03698 DMR, 2014 WL 6893866, at *21 (N.D. Cal. Dec. 5, 2014), *rev'd on other grounds*, 696 F. App'x 175 (9th Cir. 2017) (negligent and intentional interference with prospective economic advantage require the same elements, except instead of proving that the acts were intentional, plaintiffs need to prove only that the defendant was negligent).

- **Unjust enrichment.** Despite Plaintiffs' attempt to recast their unjust enrichment claim as

1   one concerning "unauthorized use of their Licensed Materials to train Codex and Copilot" (FAC ¶

2   271), the Copyright Act still preempts this claim. At its core, this claim asserts that OpenAI

3   improperly benefitted because it "train[ed] Codex and Copilot without following the licenses."

4   (*Id.* ¶ 269.) Specifically, Plaintiffs allege that (1) Defendants have "trained Codex and Copilot"

5   to "pretend[] those licenses do not exist," including because Codex "has not been trained to

6   provide Attribution"; and (2) Copilot is itself a derivative work of Plaintiffs' code.   (*Id.* ¶¶ 60,

7   181, 194 n. 34.)

8   Whether based on training or output, Plaintiffs' unjust enrichment claim therefore is one

9   for the unauthorized reproduction of Plaintiffs' code without attribution, and the unauthorized

10   creation of a derivative work. It is therefore preempted. *See Jonathan Browning, Inc. v. Venetian*

11   *Casino Resort, LLC*, No. C 07-3983JSW, 2007 WL 4532214, at *9 (N.D. Cal. Dec. 19, 2007)

12   (finding the unjust enrichment claim preempted where plaintiff argued that "the law should

13   impose an obligation on [defendant] to prevent unjust enrichment because [the parties] both

14   understood that [defendant] would not reproduce or copy the sconces unless it accepted

15   [plaintiff's] bid," as the "understanding not to reproduce or copy its sconces is 'equivalent to'"

16   the rights under the Copyright Act); *Shade v. Gorman*, No. C 08-3471 SI, 2009 WL 196400, at *5

17   (N.D. Cal. Jan. 28, 2009) (unjust enrichment claims preempted where the "gravamen of these

18   claims is that defendant surreptitiously copied plaintiff's raw footage and photographs, used those

19   materials to make [a movie] and that defendant has benefitted as a result," which were "the same

20   facts" and "same rights, as plaintiffs' copyright claim").

21   •   **Unfair Competition.** To the extent Plaintiffs' claims are based on preempted state law

22   claims, the derivative claim must also fail. *See Sulit v. Sound Choice Inc.*, No. C06-00045 MJJ,

23   2006 WL 8442163, at *7 (N.D. Cal. Nov. 14, 2006) ("State law causes of action for unfair

24   competition based on misappropriation of copyrighted material are preempted.").

25   •   **Negligence.** Plaintiffs' negligence claim is subject to preemption because allegations that

26   OpenAI breached its duty of care to comply with the open-source licenses is just a copyright

27   infringement claim by another name. The alleged duty of care is no more than a duty to refrain

28   from what Plaintiffs regard as copyright infringement. Preemption under the Copyright Act

applies when, as here, the complaint "merely recharacterizes a copyright infringement claim as one for negligence." *Dielsi v. Falk*, 916 F. Supp. 985, 992-93 (C.D. Cal. 1996). *See also AF Holdings, LLC v. Doe*, No. C 12-2049 PJH, 2012 WL 3835102, at *2 (N.D. Cal. Sept. 4, 2012) (finding preemption where plaintiff claimed negligence based on a breach of a "duty to secure [one's] Internet connection to prevent infringement of [plaintiff's] copyrighted works"); *Cromwell v. Certified Forensic Loan Auditors*, No. 17-CV-02429-DMR, 2019 WL 1095837, at *11 (N.D. Cal. Jan. 10, 2019), *report and recommendation adopted*, No. C 17-02429 SBA, 2019 WL 2181969 (N.D. Cal. Feb. 12, 2019) (dismissing negligence where plaintiffs alleged that defendant "published a copyrighted eBook on its website" but had "plead no facts for negligence separate from their copyright infringement claim"). Thus, the negligence claim is preempted by the Copyright Act.

C. Plaintiffs' Common Law Claims Also Fail Because Plaintiffs Have Not Alleged Which State's Laws Apply.

Plaintiffs' common law claims (intentional interference with prospective economic relations, negligent interference with prospective economic relations, unjust enrichment, unfair competition, and negligence) fail for the additional reason that Plaintiffs do not identify which state's law applies. This Court previously held that "due to variances among state laws, failure to allege which state law governs a common law claim is grounds for dismissal." (Order at 4 n.5 (citing *In re Nexus 6P Prods. Liab. Litig.*, 293 F. Supp. 3d 888, 933 (N.D. Cal. 2018))) (cleaned up).) Plaintiffs are citizens of New Hampshire, Illinois, Idaho, South Carolina, and Massachusetts. They assert their claims on behalf of putative nationwide classes. Yet nowhere does the FAC identify which state's (or states') laws govern Plaintiffs' claims. Plaintiffs thus fail to provide OpenAI adequate notice of their claims, as ordered by the Court. (*See* Order, at 4 n.5.)

D. Plaintiffs' Claims Fail for Reasons Specific to Each Claim.

1. Plaintiffs' DMCA Claim Should Be Dismissed.

Although Plaintiffs have dropped their DMCA § 1202(a) claim, they continue to allege that Defendants violated DMCA § 1202(b) by (1) removing or altering CMI from Licensed Materials, and (2) distributing copies of Licensed Materials knowing CMI had been removed or

1   altered without authority. (FAC ¶¶ 184-185.)

2   Plaintiffs' newly pled facts reinforce why Plaintiffs have not plausibly pled that OpenAI

3   has violated the DMCA. In their original complaint, Plaintiffs were unable to identify specific

4   instances in which Copilot output Plaintiffs' code. The FAC shows why: Because Plaintiffs

5   allege only that Copilot will output portions of their code when very specifically prompted to do

6   so, and even then those portions are modified. Even accepting Plaintiffs' allegations as true, they

7   have not pled the required elements of a claim under Section 1202(b) of the DMCA.

8   **a.   Plaintiffs Have Not Properly Pled a Claim for Removal of CMI.**

9   To plead a claim for removal of CMI, a plaintiff must plausibly allege: (1) the existence of

10   CMI conveyed in connection with a work, (2) removal or alteration of that information, (3) that

11   the removal or alteration was done intentionally; and (4) the removal or alteration was done

12   knowing or having reasonable grounds to know that it would induce, enable, facilitate, or conceal

13   copyright infringement. 17 U.S.C. § 1202(b); *Stevens v. CoreLogic, Inc.*, 899 F.3d 666, 673 (9th

14   Cir. 2018) (discussing the mental state elements); *O'Neal v. Sideshow, Inc.*, 583 F. Supp. 3d

15   1282, 1286-87 (C.D. Cal. 2022) (discussing other elements).

16   The Court found that Plaintiffs had adequately alleged removal because they had pled that

17   Defendants trained Copilot not to output CMI. (Order at 19.) But Plaintiffs have now alleged

18   specific examples of what "removal" looks like, and those examples are not enough to state a

19   claim under the DMCA. In order to get Copilot to output modified copies of portions of

20   Plaintiffs' source code, Plaintiffs had to input very specific and substantial portions of Plaintiffs'

21   source code into Copilot. For instance, the third example Plaintiffs provide is ████████████

22   ██

23   ██

24   ██

25   ██

26   ██

27   ██

28   In this example, Plaintiffs have not shown that Copilot removes CMI from "a work" at all.

1 The example shows that Plaintiffs had to provide Copilot with a significant portion of the code

2 from the work ██

3 ██

4 ██ (*Id.*) Copilot's

5 suggestion of a modified version of five lines of code does not constitute removal of CMI from

6 Doe 5's "work," when Plaintiffs themselves provided the majority of the work to Copilot (without

7 CMI) as the input.

8   The same is true of all of Plaintiffs' other examples: in each case Plaintiffs needed to

9 provide substantial material from the original in order to get Copilot to suggest a modified version

10 of the following portion of code, and in no case do any of the portions of code in question contain

11 any CMI (either the original or the code suggested by Copilot). (*See id.* ¶ 101, 106, 114, 122.)

12 Those examples render implausible the allegation that Copilot has removed CMI from their

13 works.

14   **b.   Plaintiffs Fail to Allege Removal from Identical Copies.**

15   Even if they have adequately alleged removal, Plaintiffs' claim under § 1202(b) fails for

16 the additional reason that Plaintiffs do not allege that CMI was removed from identical copies of

17 Plaintiffs' code. To prevent § 1202 from subsuming every copyright dispute, courts have

18 interpreted "removal" in the § 1202 context to require that there was some *identical* copy of the

19 plaintiff's work made without the plaintiff's CMI. *See, e.g.*, *Kelly v. Arriba Soft Corp.*, 77 F.

20 Supp. 2d 1116, 1122 (C.D. Cal. 1999) (requiring that CMI was removed from "a plaintiff's

21 product or original work"), *aff'd and rev'd in part on other grounds*, 336 F.3d 811 (9th Cir.

22 2003). Where a defendant makes a copy of a defendant's work that is substantially similar, but

23 not identical, to the plaintiff's work, and omits CMI from that copy, there may be a claim for

24 copyright infringement, but there cannot be a claim under § 1202. *See Frost-Tsuji Architects v.*

25 *Highway Inn, Inc.*, No. CIV. 13-00496 SOM, 2015 WL 263556, at *3 (D. Haw. Jan. 21, 2015),

26 *aff'd*, 700 F. App'x 674 (9th Cir. 2017) ("But the drawing by [the defendant] is not identical to

27 the drawing by [the plaintiff], such that this court can say that [the defendant] removed or altered

28 [the plaintiff's] copyright management information from [the drawing]."); *id.* ("basing a drawing

1   on [the plaintiff's] work is not sufficient to support a claim" under § 1202); *Kirk Kara Corp. v. W.*

2   *Stone & Metal Corp.*, No.CV 20-1931-DMG (EX), 2020 WL 5991503, at *6 (C.D. Cal. Aug. 14,

3   2020) (dismissing DMCA claim because "while the works may be *substantially similar*,

4   Defendant did not make *identical* copies of Plaintiff's works and then remove engraved CMI");

5   *Advanta-STAR Auto. Rsch. Corp. of Am. v. Search Optics LLC*, No. 22-CV-1186 TWR (BLM),

6   2023 WL 3366534, at *12 (S.D. Cal. May 9, 2023) (finding plaintiff had failed to state a claim for

7   a violation of § 1202 of the DMCA because plaintiff had "not plausibly alleged that Defendants

8   distributed identical copies of Plaintiff's comparison").

9   Here, Plaintiffs provide examples that affirmatively demonstrate that Copilot is *not*

10   outputting identical copies:

20   But the mere fact that the (different) code output by Copilot supposedly has the same

21   function as Plaintiffs' code isn't enough to state a DMCA claim. While the code may have the

22   same *function*, it has different expression—and it is expression that is protected by copyright law,

23   not function. *See Sony Comp. Entm't, Inc. v. Connectix Corp.*, 203 F.3d 596, 603 (9th Cir. 2000)

24   ("the Copyright Act protects expression only, not ideas or the functional aspects of a software

25   program"). Section 1202 is no different: A person must remove information conveyed "in

26   connection with *copies* or phonorecords *of a work*," 17 U.S.C. § 1202(c) (defining copyright

27   management information), not a from a different work that conveys the same functional

28

1   information.

2   Because Plaintiffs affirmatively allege that the output at issue is not identical to the

3   allegedly copied material, they have pleaded themselves out of court on the § 1202 claim, and it

4   should be dismissed with prejudice.

5   **c.   Plaintiffs Have Failed to Plead a Claim for Distributing Copies
                                   of Works from Which CMI Has Been Removed.**
6

7   Plaintiffs' claim that Defendants have distributed copies of code from which CMI has

8   been removed fails for the same reasons as its claim for removal of CMI. 17 U.S.C.

9   §§ 1202(b)(2), 1202(b)(3); *see Kirk Kara*, 2020 WL 5991503, at *6 (applying same 1202(b)(1)

10   analysis to distribution claims); *Dolls Kill, Inc. v. Zoetop Bus. Co.*, No. 2:22-cv-01463-RGK-

11   MAA, 2022 WL 16961477, at *3-4 (C.D. Cal. Aug. 25, 2022) (concluding no DMCA violation

12   for complaint that defendants "are distributing knockoff products" where the works were not

13   identical and only had "certain[] similarities"). The supposed copies are not identical.

14   **2.   Plaintiffs' Claims for Intentional and Negligent Interference with
                                   Prospective Economic Relations Fail.**
15

16   The FAC fails to state a claim for intentional or negligent interference with prospective

17   economic advantage. A cause of action for interference with prospective business advantage

18   requires: "(1) an economic relationship between the plaintiff and some third party, with the

19   probability of future economic benefit to the plaintiff; (2) the defendant's knowledge of the

20   relationship; (3) intentional acts on the part of the defendant designed to disrupt the relationship;

21   (4) actual disruption of the relationship; and (5) economic harm to the plaintiff proximately

22   caused by the acts of the defendant." *Piping Rock Partners, Inc. v. David Lerner Assocs., Inc.*,

23   946 F. Supp. 2d 957, 980 (N.D. Cal. 2013). The acts alleged must have been "wrongful by some

24   legal measure other than the fact of interference itself." *Della Penna v. Toyota Motor Sales,*

25   *U.S.A., Inc.*, 11 Cal. 4th 376, 393 (1995). A claim for negligent interference differs in that "in

26   place of the intentional conduct requirement, the plaintiff must show that the defendant owed the

27   plaintiff a duty of care which was breached by the defendant's negligent conduct." *Impeva Labs,*

28   *Inc. v. Sys. Planning Corp.*, No. 5:12-CV-00125-EJD, 2012 WL 3647716, at *6 (N.D. Cal. Aug.

1   23, 2012). As explained below, Plaintiffs' claims fail to meet the minimum pleading

2   requirements.

3   a.   **Plaintiffs Have Not Sufficiently Pled an Economic Relationship
                              with a Third Party that Has a Probability of Future Economic
4   Benefit.**

5   As an initial matter, both of Plaintiffs' interference claims fail because the amended

6   complaint does not sufficiently plead an economic relationship between Plaintiffs and a third

7   party with a probability of future economic benefit. To establish an economic relationship, "it

8   must be reasonably probable the prospective economic advantage would have been realized but

9   for defendant's interference." *Song v. Drenberg*, No. 18-CV-06283-LHK, 2019 WL 1998944, at

10   *7-8 (N.D. Cal. May 6, 2019) (cleaned up). Plaintiffs "must establish an actual economic

11   relationship or a protected expectancy with a third person, *not merely a hope of future*

12   *transactions*." *Brown v. Allstate Ins. Co.*, 17 F. Supp. 2d 1134, 1140 (S.D. Cal. 1998); *see also*

13   *Rosen v. Uber Techs., Inc.*, 164 F. Supp. 3d 1165, 1178-79 (N.D. Cal. 2016) (specific

14   relationships must be identified, rather than "hypothetical future relationship[s]").

15   Plaintiffs allege that OpenAI interfered with Plaintiffs' "prospective open-source

16   relationships" with "user communities" by "emitting code subject to open-source licenses without

17   the licenses attached," and that user communities generally create a probability of future

18   economic benefit. (FAC ¶¶ 245, 250.) But the FAC does not plead *specific* relationships

19   between Plaintiffs and third parties or existing relationships between Plaintiffs and any

20   programmers or "user communities." Plaintiffs' failure to plead specific relationships with third

21   parties is fatal to both of its interference claims. *See SunPower Corp. v. SolarCity Corp.*, No. 12-

22   CV-000694-LHK, 2012 WL 6160472, *15 (N.D. Cal. Dec. 11, 2012) (dismissing claim where

23   plaintiff had not identified specific customer relationships or facts regarding how defendant

24   interfered with those relationships); *AirDefense, Inc. v. AirTight Networks, Inc.*, No. C 05-

25   04615JF, 2006 WL 2092053, at *7 (N.D. Cal. Jul. 26, 2006) (dismissing claim where party had

26   not alleged with any specificity "with whom it had economic relationships").

27   In addition, while Plaintiffs allege that they "posted their code on GitHub with the

28   expectation that other programmers would use, modify, copy, or otherwise iterate on their posted

1   code," and that Codex and Copilot "provide an alternative interface to the same open-source

2   code," (FAC ¶¶ 247, 249), they have not pled facts demonstrating that an economic benefit would

3   have been realized but for OpenAI's alleged interference. Instead, the amended complaint

4   contains vague allegations that "users *sometimes* arrange financial contracts with authors" and

5   that "[t]he exposure from user communities *can* [] bring collateral benefits, like job offers or

6   research grants." (*Id.* ¶ 245 (emphasis added).) These statements, however, do not establish that

7   it would have been reasonably likely for *Plaintiffs* to have realized those benefits from third

8   parties. Plaintiffs' expectation of "prospective open-source relationships" (FAC ¶ 250) rests, at

9   most, on a hope of future transactions. The law precludes recovery for Plaintiffs' speculative

10   expectancies, which lack factual support demonstrating they are reasonably probable. *See*

11   *Westside Ctr. Assocs. v. Safeway Stores 23, Inc.,* 42 Cal. App. 4th 507, 527 (1996) (finding that

12   plaintiffs could not establish the requisite economic relationship because "[w]ithout an existing

13   relationship with an identifiable buyer, [plaintiff's] expectation of a future sale was 'at most a

14   hope for an economic relationship and a desire for future benefit'") (cleaned up); *Song*, 2019 WL

15   1998944, at *8 (granting motion to dismiss intentional interference with prospective business

16   relations claim, in part, because plaintiffs impermissibly rested their claim "on a hope of future

17   transactions"). Plaintiffs' interference claims should be dismissed on this ground alone.

            **b.   Plaintiffs Have Not Pled OpenAI's Knowledge of an Economic
18   Relationship with Third Parties.**

19   Both of Plaintiffs' interference claims also fail because they have not alleged that OpenAI

20   had knowledge of Plaintiffs' speculative "prospective open-source relationships." (FAC ¶ 250.)

21   To plead "knowledge by the defendant of the relationship with which the interference occurred,"

22   a defendant must know of "contracts or other 'reasonably probable' prospective economic

23   relationships." *Go Daddy Operating Co., LLC v. Ghaznavi*, No. 17-CV-06545-PJH, 2018 WL

24   1091257, at *10 (N.D. Cal. Feb. 28, 2018) (cleaned up). Here, Plaintiffs make the conclusory

25   allegation that "Defendants knew that they were interfering with Plaintiffs and Class members'

26   prospective open-source relationships." (FAC ¶ 250.) While Plaintiffs further allege that

27   "Defendants knew that Codex and Copilot were emitting code subject to open-source licenses

28

1 without the licenses attached" (*id*.), they allege no *facts* demonstrating that OpenAI knew of any

2 reasonable probable economic relationships that would be interfered with. *See Go Daddy*

3 *Operating Co., LLC*, 2018 WL 1091257, at *11 (dismissing intentional interference with

4 prospective economic advantage claim, in part, because plaintiff "does not identify any particular

5 prospective economic relationship that defendants knew about"); *see also Blazheiev v. Ubisoft*

6 *Toronto Inc.*, No. 17-cv-07160-EMC, 2018 WL 3417481, at *10 (N.D. Cal. July 13, 2018)

7 (dismissing intentional interference with prospective economic advantage claim where the

8 "Plaintiff has only made a conclusory statement in positing that 'Defendants knew of the

9 relationships[]' . . . without any factual allegations in support").

                    c.   **Plaintiffs Have Not Alleged Actual Disruption to their
10   Prospective Economic Relations.**

11   Both of Plaintiffs' interference claims also require a plausible allegation that their

12 relationships with third parties were "actually" disrupted because of OpenAI's conduct. *See*

13 *Silicon Knights, Inc. v. Crystal Dynamics, Inc.*, 983 F. Supp. 1303, 1311 (N.D. Cal. 1997).

14   Here, Plaintiffs have failed to plead facts showing actual disruption to their prospective

15 open-source relationships. The FAC's generic allegations that "Defendants intentionally

16 prevented Copilot users from becoming part of the user communities that would ordinarily

17 accrete around the open-source projects of Plaintiffs" (FAC ¶ 249) do not establish actual

18 disruption. Plaintiffs do not allege in what way or how their future relationships with specific

19 third parties were disrupted. Nor do they allege specific facts in support—for example, that they

20 lost a financial contract, job offer, or research grant because of OpenAI's actions. *See, e.g.*,

21 *Sybersound Records, Inc. v. UAV Corp.*, 517 F.3d 1137, 1151 (9th Cir. 2008) (dismissing tortious

22 interference claim because the complaint failed to allege that plaintiffs "lost a contract . . . [or]

23 that a negotiation with a Customer failed"); *Silicon Knights*, 983 F. Supp. at 1313 (finding the

24 pleadings insufficient where the complaint alleged only that the misrepresentations induced

25 distributors not to deal with plaintiffs without providing facts alleging an actual disruption to

26 negotiations or potential contracts).

27

28

1

d.   Plaintiffs Have Not Pled Economic Harm.

2

Plaintiffs make only conclusory and speculative allegations regarding the "economic

3

harm" they allegedly sustained as a result of OpenAI's purported interference. *Korea Supply Co.*

4

v. Lockheed Martin Co., 29 Cal. 4th 1134, 1153 (2003). Specifically, Plaintiffs allege that they

5

"have been deprived of the economic benefits of open-source licenses," without pleading facts

6

sufficient to show that it was reasonably probable an economic advantage would have been

7

realized. (FAC ¶ 252.) For example, Plaintiffs do not identify any prospective contracts, job

8

offers, or research assignments that they allegedly lost as a result of OpenAI's purported

9

interference. Nor do Plaintiffs set forth allegations related to their alleged monetary or

10

reputational harm. *See Go Daddy*, 2018 WL 1091257, at *11 ("general reputational harm,

11

unmoored from disrupted relationships" are insufficient to state a claim). Both of Plaintiffs'

12

interference claims fail for this additional reason.

13

e.   Plaintiffs Have Failed to Plead that OpenAI Engaged in Wrongful Conduct.

14

15

In addition, Plaintiffs have not pled that OpenAI "engaged in conduct that was wrongful

by some legal measure other than the act of interference itself." *Della Penna*, 11 Cal.4th at 393.

16

Instead, the FAC makes the conclusory allegation that Defendants collectively have "intentionally

17

and wrongfully interfered with prospective business interests and expectations of Plaintiffs."

18

(FAC ¶ 251.)   While Plaintiffs may point to their other claims against OpenAI as demonstrating

19

wrongful conduct, for the reasons stated elsewhere in this motion, those other claims fail.

20

f.   Plaintiffs Have Not Pled the Requisite Intent to Establish their Intentional Interference Claim.

21

22

To satisfy the intent element of an intentional interference with prospective economic

23

relations claim, a plaintiff must plead "acts by defendant designed to disrupt the relationship."

24

Korea Supply Co., 29 Cal. 4th at 1154. The FAC doesn't allege any *facts* regarding intentional

25

actions undertaken by OpenAI designed to disrupt Plaintiffs' prospective. Instead, Plaintiffs

26

make the conclusory allegation that "Defendants have [] intentionally and wrongfully interfered

27

with [Plaintiffs'] prospective business interests and expectations." (FAC ¶ 251.) These

28

allegations are insufficient. *See name.space, Inc. v. Internet Corp. for Assigned Names &*

Numbers, No. CV 12-8676 PA (PLAx), 2013 WL 2151478, at *8 (C.D. Cal. Mar. 4, 2013) (dismissing claim because "the Complaint does not allege any intentional actions undertaken by [defendant] designed to induce breach of Plaintiff's contracts with its clients or any evidentiary facts, as opposed to conclusory allegations . . . .") (cleaned up).

g. Plaintiffs Have Not Pled the Requisite Duty of Care Element to Establish a Negligent Interference Claim.

A claim for negligent interference with prospective economic advantage, on the other hand, "arises only when the defendant owes the plaintiff a duty of care." *Stolz v. Wong Commc'ns Ltd. P'ship*, 25 Cal. App. 4th 1811, 1825 (1994). Plaintiffs' negligent interference claim fails because the FAC does not allege that OpenAI owed Plaintiffs a duty of care. For example, Plaintiffs have not alleged that they are customers of any Defendant or that they had any relationship with any Defendant that would have created a legal duty of care. At most, the FAC alleges (albeit in connection with Plaintiffs' negligence cause of action) that OpenAI owed Plaintiffs a duty of care based on upon "contractual obligations" and "using open-source code in violation of open-source licenses to train Codex and Copilot." (FAC ¶¶ 283, 287.) However, for the reasons discussed below, that's not enough. (*See infra* IV.C.5.)

Because Plaintiffs have failed to plead any of the elements for an intentional or negligent interference claim with prospective economic relations, the claims should be dismissed.

3. Plaintiffs Fail to State a Claim for Unjust Enrichment.

Plaintiffs' claim for unjust enrichment fails because it is not an independent cause of action. "[T]here is no[] standalone cause of action for 'unjust enrichment'" under California law, and where it is alleged as such, courts instead "construe [such claims] as a quasi-contract claim for restitution." *Astiana v. Hain Celestial Grp., Inc.*, 783 F.3d 753, 762 (9th Cir. 2015); *see also Baiul-Farina v. Lemire*, 804 F. App'x 533, 537 (9th Cir. 2020) ("Unjust enrichment is not a cause of action under California law.") (cleaned up). But Plaintiffs cannot recover under a quasi-contract theory "when the parties have a valid contract regarding the same subject matter." *Hameed v. IHOP Franchising LLC*, 520 Fed. App'x 520, 522 (9th Cir. 2013) (affirming dismissal of unjust enrichment claim). While plaintiffs "may plead inconsistent claims that allege both the

1   *existence* of an enforceable agreement [for a breach of contract claim] and the *absence* of an

2   enforceable agreement" for a quasi-contract claim under the theory of unjust enrichment,

3   plaintiffs are precluded from doing so where "plaintiffs' breach of contract claim plead[s] the

4   existence of an enforceable agreement and their unjust enrichment claim d[oes] not deny the

5   existence or enforceability of that agreement." *Klein v. Chevron U.S.A., Inc.*, 202 Cal. App. 4th

6   1342, 1389 (2012). Here, Plaintiffs' claim must be dismissed because they have alleged *only* that

7   valid "contracts have been formed between Defendants on the one hand and Plaintiffs and the

8   Class on the other" based on the terms of the open-source licenses. (FAC ¶ 217.) In their unjust

9   enrichment claim, they do not "deny the existence or enforceability" of the open-source licenses.

10   (*See id.* ¶¶ 266-74.)

11   Moreover, even if unjust enrichment may be pled as a separate cause of action, Plaintiffs

12   must establish that the "defendant received and unjustly retained a benefit at the plaintiff's

13   expense." *ESG Cap. Partners, LP v. Stratos*, 828 F.3d 1023, 1038 (9th Cir. 2016). Plaintiffs rely

14   only on conclusory allegations that OpenAI "derived profit or other benefits from the use of"

15   Plaintiffs' code and that it would "be unjust for [OpenAI] to retain those benefits." (FAC ¶¶ 272-

16   73.) Plaintiffs have not alleged any *facts* to identify what the profit or benefits are, falling short

17   of their burden to plausibly allege a claim. *See Chiu v. NBS Default Servs., LLC*, No. 14-CV-

18   05261-EDL, 2015 WL 1221399, at *9 (N.D. Cal. Mar. 17, 2015) (dismissing unjust enrichment

19   claim where plaintiff's "allegations are conclusory and speculative as to how [d]efendant received

20   an unjust benefit from [p]laintiff"); *Rosal v. First Fed. Bank of Cal*. 671 F. Supp. 2d 1111, 1133

21   (N.D. Cal. 2009) (dismissing claim that "merely incorporat[ed] the other facts of the FAC by

22   reference and ma[de] a conclusory allegation that defendants have been 'unjustly enriched' by

23   'retaining profits, income and ill-gotten gains at the expense of plaintiff'").

24   **4.   Plaintiffs Fail to State an Unfair Competition Claim.**

25   Plaintiffs assert an unfair competition claim under (1) common law and (2) California's

26   UCL statute, predicated on OpenAI's alleged violations of the DMCA, violations of Plaintiffs'

27   open-source licenses, tortious interference with Plaintiffs' prospective economic advantage with

28   users of their code, and failure to attribute Codex and Copilot's output as that of Plaintiffs and the

1 purported Class. (FAC ¶ 276.) California's UCL prohibits "any unlawful, unfair or fraudulent

2 business act or practice and unfair, deceptive, untrue or misleading advertising." Cal. Bus. &

3 Prof. Code § 17200. Plaintiffs assert a claim under all three prongs of the UCL. (FAC ¶¶ 277-

4 79.) Under either the common law or the UCL, Plaintiffs' unfair competition claim fails.

5 As an initial matter, Plaintiffs' UCL claim fails because Plaintiffs have not established

6 that they lack an adequate legal remedy or that they suffered any economic injury. "Remedies

7 under the UCL are limited to restitution and injunctive relief, and do not include damages."

8 *Silvercrest Realty, Inc. v. Great Am. E&S Ins. Co.*, No. SACV 11-01197-CJC (ANx), 2012 WL

9 13028094, at *2 (C.D. Cal. Apr. 4, 2012). To state a viable claim for "equitable restitution for

10 past harm under the UCL," a plaintiff "must establish that she lacks an adequate remedy at law."

11 *Sonner v. Premier Nutrition Corp.*, 971 F.3d 834, 844 (9th Cir. 2020) (affirming dismissal where

12 plaintiff failed to allege an inadequate legal remedy). In addition, "a plaintiff bringing suit under

13 any prong of the UCL must . . . show that economic injury was the result of, i.e., caused by, the

14 unfair business practice or false advertising that is the gravamen of the claim." *Davis v.*

15 *RiverSource Life Ins. Co.*, 240 F. Supp. 3d 1011, 1017 (N.D. Cal. 2017) (cleaned up). Here,

16 Plaintiffs have not shown that no adequate legal remedy exists. Nor have Plaintiffs alleged that

17 they were economically injured as a result of OpenAI's conduct. (*See* FAC ¶ 210.) As this Court

18 previously held, Plaintiffs' UCL claim predicated on any surviving claim must fail where "no

19 such injury is alleged in the complaint." (*See* Order at 23.)

20 Plaintiffs' UCL claim under the "unlawful" prong separately fails because there is no

21 predicate violation, given that all of Plaintiffs' other claims should be dismissed for the reasons

22 explained elsewhere in this motion. When the underlying legal claim that supports a UCL claim

23 fails, "so too will the [] derivative UCL claim." *Yellowcake, Inc. v. Hyphy Music, Inc.*, No. 1:20-

24 CV-0988 AWI BAM, 2021 WL 3052535, at *13 (E.D. Cal. July 20, 2021). (*See also* Order, at

25 22-23 (citing *Eidmann v. Walgreen Co.*, 522 F. Supp. 3d 634, 647 (N.D. Cal. 2021) (If the

26 "plaintiff cannot state a claim under the predicate law … [the UCL] claim also fails.").)

27 Plaintiffs' UCL claim under the "unfair" prong also fails because the amended complaint

28 offers nothing more than a formulaic recitation of the elements of this prong. (*See* FAC ¶ 278.)

1   The amended complaint lacks allegations about the benefits of Codex and Copilot, weighed

2   against the risks of the products to the public at large, and what more OpenAI could and should

3   have done to mitigate the risks. Without any such facts, the amended complaint's bare allegations

4   are insufficient to state a claim. *See Lusinyan v. Bank of Am., N.A.*, No. CV-14-9586 DMG

5   (JCx), 2015 WL 12803453, at *2 (C.D. Cal. Sept. 15, 2015) (finding that plaintiff failed to state a

6   claim where plaintiff did not allege facts showing the consumer injury is substantial, not

7   outweighed by countervailing consumer benefits, and could not have been reasonably avoided

8   (citing *In re Sony Grant Wega KDF-E A/10/A20 Series Rear Projection HDTV Television Litig.*,

9   758 F. Supp. 2d 1077, 1091 (S.D. Cal. 2010))).

10   Plaintiffs' UCL claim under the "fraudulent" prong also fails because their allegations do

11   not satisfy the heightened pleading requirements of Federal Rule of Civil Procedure 9(b). Rule

12   9(b) requires allegations of fraud, including claims under the UCL's "fraudulent" prong, to be

13   pleaded with particularity. *Kearns v. Ford Motor Co.*, 567 F.3d 1120, 1125 (9th Cir. 2009). Rule

14   9(b) demands that the complaint identify "the who, what, when, where, and how of the

15   misconduct charged." *Cafasso, U.S. ex rel. v. Gen. Dynamics C4 Sys., Inc.*, 637 F.3d 1047, 1055

16   (9th Cir. 2011) (cleaned up). Plaintiffs allege that "consumers are likely to be deceived" because

17   "Defendants cause Codex and Copilot's output to be emitted without the proper licensing and

18   attribution required." (FAC ¶ 279.) These allegations do not satisfy Rule 9(b)'s pleading

19   standard because nowhere in the amended complaint do Plaintiffs allege that *they* were deceived

20   by Defendants and suffered injury as a result of this deception. Rather, they point to unspecified,

21   anonymous "consumers" who may have been deceived and claim that they suffered injury in the

22   form of loss of "economic benefits [associated with] the creation of open-source works." (*See id.*

23   ¶¶ 279, 281.) Moreover, to state a UCL claim under the "fraudulent" prong, Plaintiffs must also

24   allege reliance on the alleged misrepresentations. *See O'Connor v. Uber Techs., Inc.*, 58 F. Supp.

25   3d 989, 1002 (N.D. Cal. 2014) (dismissing UCL claim with prejudice because plaintiffs failed to

26   allege reliance on the alleged misrepresentation). Yet Plaintiffs do not allege that they or

27   consumers saw, much less relied on, any representation by OpenAI.

28   Lastly, Plaintiffs' common law unfair competition claim fails as Plaintiffs have not, and

OPENAI'S MOTION TO DISMISS FIRST AMENDED COMPLAINT
Case No. 4:22-cv-06823-JST
sf-5576842

1   cannot, plausibly alleged any factual basis for a false designation of origin claim. In California,

2   "[t]he common law tort of unfair competition is generally thought to be synonymous with the act

3   of 'passing off' one's goods as those of another." *Sybersound*, 517 F.3d at 1153 (cleaned up).

4   The Ninth Circuit "has consistently held that state common law claims of unfair competition and

5   actions pursuant to [the UCL] are 'substantially congruent' to [false designation of origin] claims

6   made under the Lanham Act." *Sebastian Brown Prods. LLC v. Muzooka Inc.*, No. 15-CV-01720-

7   LHK, 2016 WL 949004, at *15 (N.D. Cal. Mar. 14, 2016) (cleaned up). Plaintiffs' common law

8   unfair competition claim is nothing more than a false designation of origin claim. Plaintiffs assert

9   that Defendants have "fail[ed] to attribute Codex and Copilot's Output as originating [from]

10   Plaintiffs and the Class rather than from Copilot, GitHub, and/or OpenAI" and "pass[ed] off

11   Codex and Copilot's output without proper attribution," as well as "misappropriated and used

12   [their code] without authorization or consent to, *inter alia*, train and develop Codex and Copilot."

13   (FAC ¶¶ 276, 278, 280.) These allegations do not support any unfair competition claim here, as a

14   claim for false designation of origin must relate to the origin of tangible goods, not the authorship

15   of an intangible work like computer code. 15 U.S.C. § 1125(a)(1)(A); *Dastar Corp. v. Twentieth*

16   *Century Fox Film Corp.*, 539 U.S. 23, 37 (2003) (concluding that the phrase "origin of

17   goods…refers to the producer of the tangible goods that are offered for sale, and not to the author

18   of any idea, concept or communication embedded in those goods"); *Agence France Presse v.*

19   *Morel*, 769 F. Supp. 2d 295, 307 (S.D.N.Y. 2011) (concluding *Dastar* forecloses Lanham Act

20   claims relating to authorship). Any claim that Codex and Copilot passed off Plaintiffs' code as

21   that of OpenAI is necessarily foreclosed by these precedents, and accordingly, must be dismissed.

5.   Plaintiffs Fail to State a Claim for Negligence.

23   Plaintiffs' claim for negligence fails to plead that OpenAI owes Plaintiffs any duty, and

24   must be dismissed. To prevail on a negligence claim, a plaintiff must establish: "(1) duty; (2)

25   breach; (3) causation; and (4) damages." *Ileto v. Glock Inc.*, 349 F.3d 1191, 1203 (9th Cir. 2003).

26   "The existence of a duty of care owed by a defendant to a plaintiff is a prerequisite to establishing

27   a claim for negligence." *Langan v. United Servs. Auto. Ass'n*, 69 F. Supp. 3d 965, 987 (N.D. Cal.

28   2014). Because Plaintiffs have not alleged any recognizable duty, Plaintiffs cannot state a claim

1   for negligence.

2   Plaintiffs claim that "OpenAI owed Plaintiffs and Class members a duty of care by using

3   open-source code in violation of open-source licenses to train Codex and Copilot." (FAC ¶ 287.)

4   But Plaintiffs' remedy for any claimed breach of contract lies in contract, not in tort. While "the

5   same wrongful act may constitute both a breach of contract and [a tort], a plaintiff must still

6   identify a duty to support a claim in tort." *Green v. ADT, LLC*, No. 16-CV-02227-LB, 2016 WL

7   5339800, at *2 (N.D. Cal. Sept. 23, 2016) (citing *Erlich v. Menezes*, 21 Cal. 4th 543, 551 (1999))

8   (cleaned up). Courts generally "enforce the breach of a contractual promise through contract law,

9   except when the actions that constitute the breach violate social policy that merits the imposition

10   of tort remedies." *Id.*   Tort damages have been available for breaches of contract in limited

11   contexts: "(1) where a breach of duty directly causes physical injury; (2) for breach of the

12   covenant of good faith and fair dealing in insurance contracts; (3) for wrongful discharge in

13   violation of fundamental public policy; or (4) where the contract was fraudulently induced." *Id*.

14   (cleaned up). Under these circumstances, "the duty that gives rise to tort liability is either

15   completely independent of the contract or arises from conduct which is both intentional and

16   intended to harm." *Id.* (cleaned up). "[O]utside the insurance context, a tortious breach of

17   contract . . . may be found when (1) the breach is accompanied by a traditional common law tort,

18   such as fraud or conversion; (2) the means used to breach the contract are tortious, involving

19   deceit or undue coercion or; (3) one party intentionally breaches the contract intending or

20   knowing that such a breach will cause severe, unmitigable harm in the form of mental anguish,

21   personal hardship, or substantial consequential damages." *Id.* (cleaned up). Plaintiffs have not

22   alleged facts that would show that any of these limited circumstances apply here such that they

23   could recover tort damages for their breach of open-source license claim. *See Sustainable*

24   *Ranching Partners, Inc. v. Bering Pac. Ranches Ltd.*, No. 17-CV-02323-JST, 2017 WL 4805576,

25   at *9 (N.D. Cal. Oct. 24, 2017) (dismissing negligence claim where plaintiff did not "identif[y]

26   any allegedly tortious conduct outside of the [d]efendants' breach of the parties' contract," and

27   "the challenged conduct [for negligence] occurred within the performance of the contract").

28   Plaintiffs also generally allege that "Defendants owed a duty of reasonable care towards

1   Plaintiff and the Class based upon Defendants' relationship to them" and that "[t]his duty is based

2   upon Defendants' contractual obligations, custom and practice," and right, authority, and exercise

3   of "control over the information in its possession" as well as "the requirements of California Civil

4   Code section 1714 requiring all 'persons,' including Defendants, to act in a reasonable manner

5   towards others." (FAC ¶ 283.) But Plaintiffs have not alleged that they are customers of any

6   OpenAI entity or had any relationship with an OpenAI entity that would have created a duty.

7   (*See id.* ¶¶ 19-23.) And, as established above, Plaintiffs' duty cannot solely arise out of the

8   contractual obligations in the open-source licenses. *See Coffen v. Home Depot U.S.A. Inc.*, No.

9   16-cv-03302-PJH, 2016 WL 4719273, at *4 (N.D. Cal. Sept. 9, 2016) (dismissing negligence

10 claim with prejudice where the alleged duty "arose solely from the contractual relationship

11 between the parties"). Plaintiffs accordingly have failed to plead the existence of a duty owed by

12 any OpenAI entity.

13 **V.   CONCLUSION**

14   For all of these reasons, Plaintiffs' FAC fails to state a claim against OpenAI. The FAC

15 should be dismissed in its entirety, except for Plaintiffs' breach of contract claim.

16

17

18

19

20

21

22

23

24

25

26

27

28

1   Dated: June 29, 2023   MORRISON & FOERSTER LLP

2

3   By: /s/ Joseph C. Gratz
                                                          Joseph C. Gratz
4
                                                    MICHAEL A. JACOBS
5   MJacobs@mofo.com
                                                    JOSEPH C. GRATZ
6   JGratz@mofo.com
                                                    TIFFANY CHEUNG
7   TCheung@mofo.com
                                                    MELODY E. WONG
8   MelodyWong@mofo.com
                                                    MORRISON & FOERSTER LLP
9   425 Market Street
                                                    San Francisco, California 94105-2482
10   Telephone: (415) 268-7000
                                                    Facsimile: (415) 268-7522
11
                                                    ALLYSON R. BENNETT
12   ABennett@mofo.com
                                                    ROSE S. LEE
13   RoseLee@mofo.com
                                                    ALEXANDRA M. WARD
14   AlexandraWard@mofo.com
                                                    MORRISON & FOERSTER LLP
15   707 Wilshire Boulevard
                                                    Los Angeles, California 90017-3543
16   Telephone: (213) 892-5200
                                                    Facsimile: (213) 892-5454
17
                                                    Attorneys for Defendants OPENAI, INC., a
18   Delaware nonprofit corporation, OPENAI,
                                                    L.P., a Delaware limited partnership,
19   OPENAI OPCO, L.L.C., a Delaware limited
                                                    liability company, OPENAI GP, L.L.C., a
20   Delaware limited liability company, OPENAI
                                                    STARTUP FUND GP I, L.L.C., a Delaware
21   limited liability company, OPENAI
                                                    STARTUP FUND I, L.P., a Delaware limited
22   partnership, OPENAI STARTUP FUND
                                                    MANAGEMENT, LLC, a Delaware limited
23   liability company

24

25

26

27

28

EXHIBIT I

1

2

3

4   UNITED STATES DISTRICT COURT

5   NORTHERN DISTRICT OF CALIFORNIA

6

7   J. DOE 1, et al.,   Case No. 22-cv-06823-JST

8   Plaintiffs,

                                                ORDER GRANTING IN PART
9   v.   **DENYING IN PART MOTION TO**
                                                DISMISS
10   GITHUB, INC., et al.,
                                                Re: ECF Nos. 108, 110
11   Defendants.

12

13   Before the Court are motions to dismiss filed by Defendants GitHub, Inc., and Microsoft

14   Corporation (collectively "Defendant GitHub"), ECF No. 108; and Defendants OpenAI, Inc.,

15   OpenAI, L.P., OpenAI OPCO, L.L.C., OpenAI GP, L.L.C., OpenAI Startup Fund GP I, L.L.C.,

16   OpenAI Startup Fund I, L.P., and OpenAI Startup Fund Management, LLC (collectively

17   "Defendant OpenAI"), ECF No. 110.[1] The Court will grant the motions in part and deny them in

18   part.

19   **I.   BACKGROUND**

20   Because the facts are well-known to the parties and the Court has summarized Plaintiffs'

21   allegations in detail in its prior order, ECF No. 95, the Court will not elaborate them here.

22   Defendants previously filed motions to dismiss, which this Court granted in part and

23   denied in part. *Id.* On the question of standing, the Court agreed with Defendants that Plaintiffs

24   failed to "identify any instance of Copilot reproducing Plaintiffs' licensed code and therefore

25   failed to plead a particularized injury sufficient to confer standing [for monetary relief]." *Id.* at 7.

26   Plaintiffs did, however, establish standing for injunctive relief, as their pleadings adequately

27   _____
     [1] Although the caption of this order cites to the redacted versions of Defendants' motions to
28   dismiss, ECF Nos. 108, 110, the remainder of the order will refer to the sealed versions of these
     documents—ECF Nos. 107-3, 109-3.

1   alleged that there is "at least a substantial risk that Defendants' programs will reproduce Plaintiffs'

2   licensed code as output" in the future. *Id.* at 9. On the merits, the Court denied Defendants'

3   motions with respect to Plaintiffs' claims for breach of contract and for violations of Sections

4   1202(b)(1) and 1202(b)(3) of the Digital Millennium Copyright Act ("DMCA"). The Court did,

5   however, dismiss Plaintiffs' claims for violations of Section 1202(a) and 1202(b)(2) of the

6   DMCA, tortious interference in a contractual relationship, fraud, false designation of origin, unjust

7   enrichment, unfair competition, breach of the GitHub Privacy Policy and Terms of Service,

8   violation of the California Consumer Privacy Act ("CCPA"), and negligence with leave to amend.

9   Plaintiffs' claims for civil conspiracy and declaratory relief were dismissed with prejudice.

10   In their first amended complaint ("FAC"), Plaintiffs have added a fifth Doe GitHub user as

11   a plaintiff and bring eight claims for relief: (1) violation of Sections 1202(b)(1) and 1202(b)(3) of

12   the DMCA, 17 U.S.C. §§ 1201–05; (2) common law breach of contract for open-source license

13   violations; (3) common law breach of contract for selling licensed materials;[2] (4) common law

14   intentional interference with prospective economic relations; (5) common law negligent

15   interference with prospective economic relations; (6) common law unjust enrichment; (7) common

16   law unfair competition in violation of Cal. Bus. & Prof. Code §§ 17200, *et seq.*; and (8) common

17   law negligence.

18   In support of these claims, Plaintiffs renew their allegations that Defendants "published

19   Licensed Materials [that Plaintiffs] owned a copyright interest in to at least one GitHub

20   repository" without proper copyright management information ("CMI"). ECF No. 97-3 ¶ 19; *see*

21   *id.* ¶¶ 20–23, 191. Plaintiffs also assert that "[t]hough Output from Copilot is often a verbatim

22   copy, even more often it is a modification," meaning "a near-identical copy that contains only

23   semantically insignificant variations of the original Licensed Materials, or a modified copy that

24   recreates the same algorithm." *Id.* ¶ 96.

25   Following this Court's finding that Plaintiffs lacked standing for monetary relief, the FAC

26   now includes allegations that Defendants' programs released, or "output," code published to

27

28   [2] Plaintiffs' claim for breach of contract for selling licensed materials is alleged only against
     Defendant GitHub, Inc. ECF No. 97-3 at 60.

United States District Court
Northern District of California

1   GitHub by Does 1, 2, and 5.

2   Beginning with Doe 1, "[o]n May 24, 2020, Doe 1 posted a Go source file on GitHub

3   called board.go subject to the MIT License." *Id.* ¶ 106. This code "sets up a chess board with its

4   pieces" and includes two functions: "setPieces" and "ResetBoard." *Id.* According to the FAC,

5   "[w]hen Copilot is prompted with the setPieces function and the beginning of the ResetBoard

6   function," "[t]he first suggestion from Copilot is a modification of Doe 1's code." *Id.* ¶ 107.

7   Further, the FAC alleges that "[t]he text strings 'board.setPieces(nearColor' and

8   'board.setPieces(farColor' do not appear in any other source file on GitHub[,]" and thus, "[t]he

9   only way Copilot knows how to make this suggestion is because it ingested Doe 1's source file as

10 training data." *Id.* ¶ 108. In light of this, the FAC avers that "the Copilot suggestion needs to

11 follow the requirements of Doe 1's license for that code, including providing attribution[,]" and it

12 currently does not. *Id.* ¶ 112.

13   Turning to Doe 2, the FAC alleges that "[o]n November 1, 2019, Doe 2 posted a Java

14 source file on GitHub called AminoAcid.java subject to the GNU General Public License v3.0.

15 The code contains an enumeration of the 20 amino acids and their codons." *Id.* ¶ 101. When

16 Copilot was prompted with the first few lines of Doe 2's code, including the function name

17 "AminoAcid," and the first amino acid and codons in Doe 2's code, its output was "identical to

18 Doe 2's code," save for a few "cosmetic" differences in word choice. *Id.* ¶ 103. According to the

19 FAC, "the Copilot suggestion is a nearly verbatim reproduction of Doe 2's unique code," and thus

20 "it follows that Copilot copied Doe 2's code." *Id.* ¶ 104.

21   Finally, the FAC includes descriptions of two sets of Doe 5's code posted on GitHub:

22 "hostname_test.py" and "js_to_hid_test.py." *Id.* ¶¶ 114, 122. The first code set,

23 "hostname_test.py," contains three tests "for other code that validates network hostnames." *Id.*

24 ¶ 114. Those three tests include a test to accept hostnames with valid characters ("valid characters

25 test"), a test to accept hostnames with 63 characters ("63-characters test"), and a test to reject

26 hostnames that are not a string ("non-string test"). *Id.* When a user prompts Copilot "with the

27 first section of Doe 5's code, comprising the first complete test and the name of the second[,]"

28 "Copilot offers to complete the prompt with a verbatim copy of Doe 5's original code, except that

3

1 the variable hostname_63_chars has been renamed hostname_valid (a variation that does not affect

2 how the code works)." *Id.* ¶¶ 115–16.

3   Doe 5's second set of code, "js_to_hid_test.py," likewise contains three tests for different

4 keystrokes: simple keystroke, shifted keystroke, and keystroke with all modifiers. *Id.* ¶ 122.

5 When a user prompts Copilot "with the first section of Doe 5's code, comprising the first complete

6 test and the name of the second," "Copilot offers to complete the second test with a verbatim copy

7 of Doe 5's original code." *Id.* ¶¶ 122–23.

8   Defendants now move to dismiss six of the eight claims raised in the FAC pursuant to

9 Rules 12(b)(1) and 12(b)(6) of the Federal Rules of Civil Procedure.[3] ECF Nos. 107-3, 109-3.

10 They argue that Plaintiffs lack standing to pursue claims for damages, and that Plaintiffs' amended

11 complaint fails to state a claim.

12 **II.   JURISDICTION**

13   The Court has jurisdiction over Plaintiffs' federal claims pursuant to 28 U.S.C. § 1331 and

14 supplemental jurisdiction over Plaintiffs' state law claims under 28 U.S.C. § 1367.

15 **III.   LEGAL STANDARD**

16   **A.   Rule 12(b)(1)**

17   "Article III of the Constitution confines the federal judicial power to the resolution of

18 'Cases' and 'Controversies.'" *TransUnion LLC v. Ramirez*, 594 U.S. 413, 423 (2021). "No case

19 or controversy exists if a plaintiff lacks standing or if a case is not ripe for adjudication, and

20 consequently a federal court lacks subject matter jurisdiction." *Temple v. Abercrombie*, 903 F.

21 Supp. 2d 1024, 1030 (D. Haw. 2012) (citations and quotations omitted). A defendant may attack a

22 plaintiff's lack of standing jurisdiction by moving to dismiss for lack of jurisdiction under Rule

23 12(b)(1) of the Federal Rules of Civil Procedure. *See Cetacean Cmty. v. Bush*, 386 F.3d 1169,

24 1174 (9th Cir. 2004); *see also Maya v. Centex Corp.*, 658 F.3d 1060, 1067 (9th Cir. 2011)

25 ("[L]ack of Article III standing requires dismissal for lack of subject matter jurisdiction under

26 Federal Rule of Civil Procedure 12(b)(1).").

27

28 [3] Defendants do not move to dismiss Plaintiffs' claims for breach of contract for open-source
license violations or breach of contract for selling licensed materials. *See supra* at 2.

4

1   "A Rule 12(b)(1) jurisdictional attack may be facial or factual." *Safe Air for Everyone v.*

2   *Meyer*, 373 F.3d 1035, 1039 (9th Cir. 2004). "In a facial attack, the challenger asserts that the

3   allegations contained in a complaint are insufficient on their face to invoke federal jurisdiction."

4   *Id.* Where, as here, a defendant makes a facial attack, the court assumes that the complaint's

5   allegations are true and draws all reasonable inferences in the plaintiff's favor. *See Wolfe v.*

6   *Strankman*, 392 F.3d 358, 362 (9th Cir. 2004).

7   **B.   Rule 12(b)(6)**

8   "Dismissal under Rule 12(b)(6) is appropriate only where the complaint lacks a cognizable

9   legal theory or sufficient facts to support a cognizable legal theory." *Mendiondo v. Centinela*

10   *Hosp. Med. Ctr.*, 521 F.3d 1097, 1104 (9th Cir. 2008). To survive a motion to dismiss, "a

11   complaint must contain sufficient factual matter, accepted as true, to 'state a claim to relief that is

12   plausible on its face.'" *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) (quoting *Bell Atlantic Corp. v.*

13   *Twombly*, 550 U.S. 544, 570 (2007)). "A claim has facial plausibility when the plaintiff pleads

14   factual content that allows the court to draw the reasonable inference that the defendant is liable

15   for the misconduct alleged." *Id.* In determining whether a plaintiff has met the plausibility

16   requirement, a court must "accept all factual allegations in the complaint as true and construe the

17   pleadings in the light most favorable to the nonmoving party." *Knievel v. ESPN*, 393 F.3d 1068,

18   1072 (9th Cir. 2005).

19   **C.   Leave to Amend**

20   Leave to amend a complaint "shall be freely given when justice so requires." Fed. R. Civ.

21   P. 15(a)(2). The decision of whether to grant leave to amend is "within the discretion of the

22   district court, which may deny leave due to 'undue delay, bad faith or dilatory motive on the part

23   of the movant, repeated failure to cure deficiencies by amendments previously allowed, undue

24   prejudice to the opposing party by virtue of allowance of the amendment, and futility of

25   amendment.'" *Leadsinger, Inc. v. BMG Music Pub.*, 512 F.3d 522, 532 (9th Cir. 2008) (quoting

26   *Foman v. Davis*, 371 U.S. 178, 182 (1962)).

27

28

United States District Court
Northern District of California

1   **IV.   DISCUSSION**

2   **A.   Article III Standing**

3   In its order resolving prior motions to dismiss, the Court found that Plaintiffs had standing

4   to seek injunctive relief, as they plausibly alleged that there is "at least a substantial risk that

5   Defendants' programs will reproduce Plaintiffs' licensed code as output" in the future. ECF No.

6   95 at 9. Plaintiffs, however, failed to establish standing to seek retrospective relief for money

7   damages because Plaintiffs did not allege "that they themselves have suffered the injury they

8   describe[d]." *Id.* at 8.

9   Plaintiffs' first amended complaint largely resembles their initial version with one

10   significant addition—they now include examples of licensed code owned by Does 1, 2, and 5 that

11   has been output by Copilot. ECF No. 97-3 ¶¶ 97–128. Plaintiffs aver that these allegations

12   establish a "particular personalized injury" sufficient to confer standing for damages, as such facts

13   demonstrate that Defendants removed their CMI and emitted their code in violation of their open-

14   source licenses. ECF No. 140 at 12. Defendants contend that even with these new allegations,

15   Plaintiffs still fall short of demonstrating standing for money damages. In their view, Does 1, 2,

16   and 5 have "manufactured injury to establish evidence of past harm," ECF No. 147 at 8, while

17   Does 3 and 4 have again failed to raise instances in which their code was output by Copilot. ECF

18   No. 109-3 at 18. Although the Court agrees with Defendants that Does 3 and 4 have not

19   established standing for monetary damages, it finds that Does 1, 2, and 5 have adequately alleged

20   "particular personalized injury" sufficient to confer standing for monetary damages.

21   To establish standing, "a plaintiff must show (i) that he suffered an injury in fact that is

22   concrete, particularized, and actual or imminent; (ii) that the injury was likely caused by the

23   defendant; and (iii) that the injury would likely be redressed by judicial relief." *TransUnion*, 594

24   U.S. at 423. A plaintiff bears the burden of establishing standing, but that burden is low at the

25   pleading stage, where "general factual allegations of injury resulting from the defendant's conduct

26   may suffice[.]" *Lujan v. Defs. of Wildlife*, 504 U.S. 555, 561 (1992). Moreover, "standing is not

27   dispensed in gross; rather, plaintiffs must demonstrate standing for each claim that they press and

28   for each form of relief that they seek (for example, injunctive relief and damages)." *TransUnion*,

1   594 U.S. at 431.

2   In a putative class action, named plaintiffs "must allege and show that they personally have

3   been injured, not that injury has been suffered by other, unidentified members of the class to

4   which they belong and which they purport to represent." *Warth v. Seldin*, 422 U.S. 490, 502

5   (1975).

6   **a.   Does 1, 2, and 5**

7   Relying on *Clapper v. Amnesty Int'l USA*, 568 U.S. 398, 413 (2013), Defendants contend

8   that Does 1, 2, and 5 have not adequately alleged past harm because "a plaintiff's own acts"—*i.e.*,

9   inputting their own code into Copilot to demonstrate output—"cannot give rise to a 'concrete

10   injury.'" ECF No. 147 at 8. In *Clapper*, plaintiffs challenged the constitutionality of Section 702

11   of the Foreign Intelligence Surveillance Act of 1978, which "allows the Attorney General and the

12   Director of National Intelligence to acquire foreign intelligence information by jointly authorizing

13   the surveillance of individuals who are not 'United States persons' and are reasonably believed to

14   be located outside the United States." *Clapper*, 568 U.S. 398 at 401. Plaintiffs were "United

15   States persons whose work [allegedly] require[d] them to engage in sensitive international

16   communications with individuals who they believe[d] [were] likely targets of surveillance under"

17   that statute. *Id.* As a basis for standing, plaintiffs alleged that, in response to the surveillance

18   statutes, they had to incur costly and burdensome measures to protect the confidentiality of their

19   communications. *Id.* at 401–02. The Supreme Court held that a plaintiff "cannot manufacture

20   standing by choosing to make expenditures based on hypothetical future harm that is not certainly

21   impending." *Id.* at 402.

22   This case is not *Clapper*. The Court already determined in its prior order that Plaintiffs

23   face "at least a substantial risk that Defendants' programs will reproduce Plaintiffs' licensed code

24   as output" in the future. ECF No. 95 at 9. The present inquiry concerns whether Plaintiffs have

25   alleged "that they have suffered a qualifying injury-in-fact that has actually occurred to them[.]"

26   *Id*. To make such a showing, a plaintiff is not required to suffer an injury only inadvertently. *See,*

27   *e.g.*, *Fed. Election Comm'n v. Cruz*, 596 U.S. 289, 297 (2022). Indeed, the Supreme Court has

28   made clear that an injury resulting from an unlawful act "remains fairly traceable" to that act,

1   "even if the injury could be described in some sense as willingly incurred." *Id.* (holding that

2   plaintiff suffered injury even though plaintiff willingly incurred a statutory penalty; distinguishing

3   *Clapper*); *see also Havens Realty Corp. v. Coleman*, 455 U.S. 363, 374–75 (1982) (finding

4   plaintiff "has standing to sue in his capacity as a tester."). Plaintiffs submit that their code was

5   emitted by Defendants' programs without their CMI in violation of their open-source licenses.

6   ECF No. 97-3 ¶¶ 19–23, 191. That Plaintiffs themselves input their code into Defendants'

7   programs (which caused the output) does not render their injury non-concrete.[4] Accordingly,

8   Plaintiffs have made "a factual showing of perceptible harm" to establish standing for monetary

9   damages. *Lujan*, 504 U.S. at 566.

10   In support of their position, Defendants contend that Plaintiffs "have not alleged any facts

11   giving reason to believe that a real-world user plausibly has or would enter the sorts of prompts

12   Plaintiffs used in their examples." ECF No. 146 at 8. They assert that Plaintiffs have neither

13   explained that their code "frequently recurs in GitHub repositories," nor that "anyone would want

14   to copy their code." *Id.* at 10. Maybe so, but Article III does not impose such requirements to

15   confer standing for monetary damages. Further, the amount of damages for past harm suffered is a

16   separate inquiry from whether Plaintiffs have alleged standing for damages in the first place.

17   Defendants' final argument is that Plaintiffs have not alleged that Copilot generated the

18   output in question prior to the commencement of this action. ECF Nos. 107-3 at 18, 109-3 at 18.

19   Defendants are correct that "[t]he existence of federal jurisdiction ordinarily depends on the facts

20   as they exist when the complaint is filed." *Lujan*, 504 U.S. at 569 n.4 (quoting *Newman–Green,*

21   *Inc. v. Alfonzo–Larrain*, 490 U.S. 826, 830 (1989)) (emphasis omitted). But "when a plaintiff files

22   an amended complaint, the amended complaint supersedes the original, the latter being treated

23   thereafter as non-existent." *Rhodes v. Robinson*, 621 F.3d 1002, 1005 (9th Cir. 2010) (citations

24   and internal quotations omitted). Plaintiffs generated this output before filing their first amended

25   complaint, which is now the operative complaint in this litigation. The Court finds Plaintiffs'

26   allegations sufficient to confer standing for monetary damages.

27

28

---

[4] Nor does it render their injury non-traceable, as Defendant GitHub alleges. *See* ECF No. 107-3 at 17. Plaintiffs' asserted injury was not only the consequence of their own actions, but rather derived from Defendants' programs. *See Warth*, 422 U.S. at 505.

b.   Does 3 and 4

Turning to Does 3 and 4, the Court agrees with Defendants that they have "yet again failed to plead specific instances in which *their* code was output by Copilot." ECF No. 109-3 at 18. "[A]t an irreducible minimum, [Article] III requires the party who invokes the court's authority to 'show that he personally has suffered some actual or threatened injury.'" *Valley Forge Christian Coll. v. Ams. United for Separation of Church and State, Inc.*, 454 U.S. 464, 472 (1982) (quoting *Gladstone Realtors v. Village of Bellwood*, 441 U.S. 91, 99 (1979)). Because Does 3 and 4 have not alleged instances where their code has been output, they have not demonstrated "a qualifying injury-in-fact that has actually occurred to them." ECF No. 95 at 9. Having previously found that Does 3 and 4 failed to allege standing for monetary damages on this same ground, the Court will now dismiss their request for monetary damages with prejudice.

In sum, the Court finds that Does 1, 2, and 5 have standing to pursue claims for both injunctive relief and damages, whereas Does 3 and 4 have standing to pursue only claims for injunctive relief.

B.   Rule 12(b)(6)

Defendants move to dismiss most of Plaintiffs' claims for failure to state a claim on which relief may be granted. The Court grants Defendants' motions with regard to their copyright preemption claims, as well as their claims under Sections 1202(b)(1) and 1202(b)(3) of the DMCA.

1.   Copyright Preemption

Defendants argue that Plaintiffs' state law claims—including intentional and negligent interference with prospective economic relations, unjust enrichment, unfair competition, and negligence—are preempted by Section 301 of the Copyright Act. For reasons set forth below, the Court agrees.

The Copyright Act of 1976 expressly preempts state law claims where the plaintiff's work "come[s] within the subject matter of copyright" and the state law grants "legal or equitable rights that are equivalent to any of the exclusive rights within the general scope of copyright[.]" 17 U.S.C. § 301(a). The rights protected under the Copyright Act include the rights of reproduction,

1   preparation of derivative works, distribution, performance, and display. 17 U.S.C. § 106.

2   The Ninth Circuit has established a two-part test to determine whether state law claims are

3   preempted by the copyright law. First, the court decides "whether the 'subject matter' of the state

4   law claim falls within the subject matter of copyright as described in 17 U.S.C. §§ 102 and 103."

5   *Laws v. Sony Music Ent., Inc*., 448 F.3d 1134, 1137 (9th Cir. 2006). If it does, the court must then

6   "determine whether the rights asserted under state law are equivalent to the rights contained in 17

7   U.S.C. § 106[.]" *Id.* at 1138. On the other hand, "[i]f a state law claim includes an 'extra element'

8   that makes the right asserted qualitatively different from those protected under the Copyright Act,

9   the state law is not preempted by the Copyright Act." *Altera Corp. v. Clear Logic, Inc.*, 424 F.3d

10   1079, 1089 (9th Cir. 2005) (quoting *Summit Mach. Tool Mfg. v. Victor CNC Sys.*, 7 F.3d 1434,

11   1439–40 (9th Cir. 1993)).

12   As a threshold matter, Plaintiffs do not appear seriously to dispute that the "subject matter"

13   of their state law claims falls within the subject matter of copyright. Copyright encompasses

14   "original works of authorship" including literary, musical, or dramatic works (among other

15   categories) that are "fixed in any tangible medium of expression . . . from which they can be

16   perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or

17   device." 17 U.S.C. § 102(a). Plaintiffs' claims concern computer code, which is a form of

18   "literary work[]" under Section 102(a). *See Google LLC v. Oracle Am., Inc*., 593 U.S. ----, 141 S.

19   Ct. 1183, 1198 (2021). Thus, the Court finds the first prong of the preemption test satisfied.

20   The dispositive question, accordingly, is whether Plaintiffs' state law claims assert

21   "equivalent rights" to those protected under Section 106 of the Copyright Act. Defendants

22   contend that although Plaintiffs label their claims as "unauthorized use," the gravamen of their

23   state law claims sound in rights equivalent to those protected under Section 106 and are thus

24   preempted. *See* ECF Nos. 107-3 at 22–24, 109-3 at 19–21. Relying on this Court's prior order,

25   Plaintiffs respond that "claims concerning unauthorized use are not preempted by the Copyright

26   Act." ECF No. 140 at 23; *see Altera*, 424 F.3d at 1089–90.

27   **a.**   **Intentional and Negligent Interference with Prospective Economic Relations**

28

10

1   The Court begins with Defendants' argument that Plaintiffs' state law claims for

2 intentional and negligent interference with prospective economic relations are preempted.[5]

3 Plaintiffs contend that "[w]hen a programmer uses an open-source software, a contract is formed

4 based on the terms of the particular open-source license" and when Defendants "emit[] code

5 subject to open-source licenses without the licenses attached," they are interfering with Plaintiffs'

6 "prospective open-source relationship." ECF No. 97-3 ¶¶ 242, 250–51. Despite its stylization, the

7 core of this claim is that "Plaintiffs have a tort-based exclusive right to control the reproduction

8 and distribution of certain code to the general public[.]" ECF No. 107-3 at 23. Accordingly, "the

9 right [Plaintiffs] seek to protect is coextensive with [] exclusive right[s] already safeguarded by the

10 Act," including the rights of reproduction, distribution, and preparation of derivative works of

11 copyrighted material. *Harper & Row Publishers, Inc. v. Nation Enterprises*, 723 F.2d 195, 201

12 (2d Cir. 1983), *rev'd on other grounds*, 471 U.S. 539, 105 (1985); *see also Media.net Advert. FZ-*

13 *LLC*, 156 F. Supp. 3d at 1073 (holding intentional interference with prospective economic

14 relations claim preempted because it was "predicated on Defendant's unauthorized copying of

15 Plaintiff's HTML code, in violation of its statutory rights under the Copyright Act.").[6]

16   **b.   Unjust Enrichment**

17   Defendants similarly argue that Plaintiffs' state law claim for unjust enrichment is

18 preempted by the Copyright Act. This Court previously held that Plaintiffs' unjust enrichment

---

19   [5] To establish a claim of intentional interference with prospective economic advantage, a plaintiff
20 must show: "(1) an economic relationship between the plaintiff and a third party, with a
probability of future economic benefit to the plaintiff; (2) the defendant's knowledge of this
21 relationship; (3) intentional and wrongful conduct on the part of the defendant, designed to
interfere with or disrupt the relationship; (4) actual disruption or interference; and (5) economic
22 harm [to] the plaintiff as a proximate result of the defendant's wrongful conduct." *Media.net
Advert. FZ-LLC v. NetSeer, Inc.*, 156 F. Supp. 3d 1052, 1073 (N.D. Cal. 2016). Similarly, to
23 prevail on a claim for negligent interference with prospective economic advantage, courts consider
the following elements: "(1) the extent to which the transaction was intended to affect the plaintiff,
24 (2) the foreseeability of harm to the plaintiff, (3) the degree of certainty that the plaintiff suffered
injury, (4) the closeness of the connection between the defendant's conduct and the injury
25 suffered, (5) the moral blame attached to the defendant's conduct, and (6) the policy of preventing
future harm." *Hsu v. OZ Optics Ltd.*, 211 F.R.D. 615, 621 (N.D. Cal. 2002).
26 [6] Plaintiffs do not argue that an extra element qualitatively changes the nature of their claims. But
even if they did, courts have already rejected the argument that a loss of business stemming from
27 the defendant's sale of plaintiff's copyrighted work constitutes an extra element to an intentional
interference with prospective economic relations claim because "federal copyright laws already
28 protect the exclusive right of distribution." *Media.net Advert. FZ-LLC*, 156 F. Supp. 3d at 1073–
74 (quoting *Aagard v. Palomar Builders, Inc.*, 344 F. Supp. 2d 1211, 1219 (E.D. Cal. 2004)).

1   claim, which alleged that Defendants unlawfully reproduced Plaintiffs' code as output and

2   prepared derivative works, was subject to preemption because the claim was based upon rights

3   "protected by the federal Copyright Act."[7] ECF No. 95 at 17 (quoting *Altera*, 424 F.3d at 1079).

4   Plaintiffs now bring a modified unjust enrichment claim, asserting that "Defendants used

5   Plaintiffs' Licensed Materials to train Codex and Copilot without following the licenses under

6   which the Licensed Materials were published." ECF No. 97-3 ¶ 269. Defendants contend that it

7   is inconsequential whether Plaintiffs' claim is based on unlawful use of "output" or "training

8   data"—both claims fall subject to preemption. *See* ECF Nos. 107-3 at 24, 109-3 at 20.

9   In support of their position, Plaintiffs rely on *Altera*, which held that claims "concerning

10   the unauthorized use of the software's end-product is not within the rights protected by the federal

11   Copyright Act." *Altera*, 424 F.3d at 1090; *see* ECF No. 140 at 23. Plaintiffs' selective reliance on

12   *Altera* is misplaced. *Altera* concerned a licensing dispute between Altera and Clear Logic, two

13   competitors in the semiconductor industry. *Altera*, 424 F.3d at 1081. Pursuant to a licensing

14   agreement, Altera customers were permitted to use Altera's software, which programmed

15   semiconductor chips, for the "sole purpose of programming [chips] manufactured by . . . and sold

16   by [Altera]." *Id.* at 1090. Clear Logic's business model relied on Altera customers using Altera

17   software to create a file called a bitstream, which Clear Logic then used to create "a different type

18   of chip" for the customer. *Id.* at 1082. Altera brought state law claims against ClearLogic "for

19   inducing Altera's customers to intentionally breach their license agreements with Altera and also

20   for intentionally interfering with those contractual relations." *Id*. Clear Logic contended that

21   those claims were preempted by federal copyright law. *Id*. In finding that the claims were not

22   preempted, the *Altera* court held that "[t]he right at issue is not the reproduction of the software as

23   Clear Logic argues, but is more appropriately characterized as the use of the bitstream." *Id*. at

24   1089.

25   While the *Altera* dispute concerned the "unauthorized use of the software's end-product,"

26   *id.* at 1090, Plaintiffs' claims principally concern the unauthorized reproduction of their code to

27   prepare derivative works—not the unlawful use of an end-product or output. Such claims fall

28   _____

[7] Plaintiffs' claim for unjust enrichment was dismissed with leave to amend. ECF No. 95 at 18.

1   under the purview of the Copyright Act. *Cf. G.S. Rasmussen & Assoc. v. Kalitta Flying Service*,

2   958 F.2d 896, 904 (9th Cir. 1992), *cert. denied*, 508 U.S. 959 (1993) ("Copyright preemption is

3   both explicit and broad"). Indeed, Plaintiffs' complaint explicitly alleges that "Defendants had

4   access to but were not licensed by Plaintiffs . . . to create Derivative Works based upon the

5   Licensed Materials." ECF No. 97-3 ¶ 194. Thus, the Court finds that Plaintiffs' unjust

6   enrichment claim is preempted.

7   c.   **Negligence**

8   The Court turns next to Defendants' argument that Plaintiffs' state law claim for

9   negligence is subject to preemption. In their FAC, Plaintiffs assert that Defendants owed them a

10 duty of reasonable care and breached this duty by negligently "engineering, designing,

11 maintaining, and controlling systems—including Codex and Copilot—which are trained on

12 Plaintiffs' . . . Licensed Materials without their authorization." *Id.* ¶ 284. Defendants respond that

13 "[t]he alleged duty of care is no more than a duty to refrain from what Plaintiffs regard as

14 copyright infringement." ECF No. 109-3 at 20. It is well-established that where "the essential

15 allegation" of a negligence claim is that a defendant "unlawfully copied" a plaintiff's idea, "it is

16 still a copyright infringement claim." *Dielsi v. Falk*, 916 F. Supp. 985, 992 (C.D. Cal. 1996).

17 Here, because Plaintiffs "merely recharacterize[] a copyright infringement claim as one for

18 negligence," the Court finds that it too is preempted by the Copyright Act. *Id.*

19   d.   **Unfair Competition**

20   Finally, Defendants argue that to the extent that Plaintiffs' unfair competition claim is

21 predicated on their state law claims for intentional and negligent interference with prospective

22 economic relations, unjust enrichment, and negligence, it must also be preempted. The Court

23 agrees. When the underlying claims are dismissed on preemption grounds, a UCL claim

24 predicated on the same causes of action must be dismissed as well. *See Kodadek v. MTV*

25 *Networks, Inc.*, 152 F.3d 1209, 1213 (9th Cir. 1998) (finding plaintiff's UCL claim preempted

26 where it was "based solely on rights equivalent to those protected by the federal copyright laws.");

27 *Maloney v. T3Media, Inc.*, 853 F.3d 1004, 1019 (9th Cir. 2017) (same).

28   Having previously dismissed Plaintiffs' unjust enrichment claim on preemption grounds,

United States District Court
Northern District of California

13

1   the Court now dismisses this claim with prejudice. Plaintiffs' unfair competition claim is
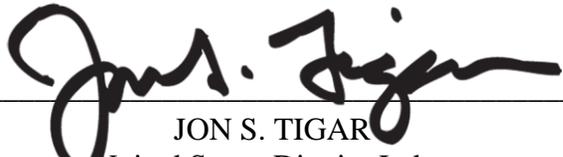
2   dismissed to the extent it is predicated on their state law claims for intentional and negligent

3   interference with prospective economic relations, unjust enrichment, and negligence. Finally,

4   because the Court is unconvinced that Plaintiffs can cure their intentional and negligent

5   interference with prospective economic relations and negligence claims, the Court likewise

6   dismisses them with prejudice.

7   **2.   DMCA Section 1202(b)(1) and 1202(b)(3)**

8   In its prior order, the Court denied Defendants' motions to dismiss Plaintiffs' claims under

9   Sections 1202(b)(1) and 1202(b)(3) of the DMCA, reasoning that Plaintiffs pleaded "sufficient

10   facts to support a reasonable inference that Defendants intentionally designed the programs to

11   remove CMI from any licensed code they reproduce as output." ECF No. 95 at 19. Further, it

12   found that Plaintiffs' allegations plausibly suggested that "Defendants knew or had reasonable

13   grounds to know that removal of CMI carried a substantial risk of inducing infringement." *Id.* at

14   20.

15   Defendants now ask the Court to address an unresolved argument from the prior briefing—

16   namely, that "[Section] 1202(b) claims lie only when CMI is removed or altered from an *identical*

17   copy of a copyrighted work." ECF No. 107-3 at 20 (emphasis added); *see* ECF No. 109-3 at 23–

18   24. Defendants argue that because Plaintiffs' new allegations state that output from Copilot is

19   often a modification of their licensed works, as opposed to an "identical copy," they have

20   effectively pleaded themselves out of their Section 1202(b)(1) and 1202(b)(3) claims. ECF No.

21   109-3 at 23. Agreeing with Defendants on both fronts, the Court finds that it is not precluded

22   from analyzing this claim anew and that Section 1202(b) claims require that copies be "identical."[8]

23   CMI includes "information such as the title, the author, the copyright owner, the terms and

24   conditions for use of the work, and other identifying information set forth in a copyright notice or

25   conveyed in connection with the work." *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 671 (9th Cir.

26   2018). To state a claim under Section 1202(b)(1), Plaintiffs must plausibly allege that Defendants

27   ───────────────────────

28   [8] A court may consider a motion "to the extent it presents issues not previously resolved." *Jones v. Life Ins. Co. of N. Am.*, No. 08-CV-03971-RMW, 2015 WL 8753996, at *3 (N.D. Cal. Dec. 15, 2015).

United States District Court
Northern District of California

1   (1) "intentionally remov[ed] or alter[ed]" CMI while "knowing, or . . . having reasonable grounds

2   to know, that it will induce, enable, facilitate, or conceal an infringement of any right under this

3   title." 17 U.S.C. § 1202(b)(1). Similarly, a violation of Section 1202(b)(3) requires that Plaintiffs

4   plausibly allege that Defendants "distribute" or "import for distribution" copies of works

5   "knowing that [CMI] has been removed or altered without authority of the copyright

6   owner[.]" 17 U.S.C. § 1202(b)(3); *see also Stevens*, 899 F.3d at 674 (requiring plaintiff to

7   demonstrate "pattern of conduct or modus operandi" to establish the requisite mental state);

8   *Falkner v. Gen. Motors LLC*, 393 F. Supp. 3d 927, 938 (C.D. Cal. 2018).

9   "Courts have held that no DMCA violation exists where the works are not identical."

10   *Advanta-STAR Auto. Rsch. Corp. of Am. v. Search Optics, LLC*, No. 22-CV-1186 TWR (BLM),

11   2023 WL 3366534, at *12 (S.D. Cal. May 9, 2023) (internal quotations and citations omitted).

12   "[E]ven where the underlying works are similar, courts have found that no DMCA violation

13   exists" unless the works are identical. *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-

14   1931-DMG, 2020 WL 5991503, at *6 (C.D. Cal. Aug. 14, 2020); *see also Frost-Tsuji Architects v.*

15   *Highway Inn, Inc.*, No. CIV. 13-00496 SOM, 2015 WL 263556, at *3 (D. Haw. Jan. 21, 2015),

16   *aff'd*, 700 F. App'x 674 (9th Cir. 2017) (finding no Section 1202(b) violation where the allegedly

17   infringing drawing was "not identical."). Plaintiffs' amended complaint alleges that "[t]hough

18   Output from Copilot is often a verbatim copy, even more often it is a modification: for instance, a

19   near-identical copy that contains only semantically insignificant variations of the original Licensed

20   Materials, or a modified copy that recreates the same algorithm." ECF No. 97-3 ¶ 96. Indeed, the

21   examples Plaintiffs provide with respect to Does 1, 2, and 5 state that the Copilot output is a

22   "modified format," "variation[]," or the "functional[] equivalent" of the licensed code. *Id.* ¶¶ 103,

23   110, 120. This, however, is not sufficient for a Section 1202(b) claim. Accordingly, the Court

24   agrees with Defendants that this is a "fundamental defect" "endemic to Plaintiffs' theory of

25   [Section] 1202(b) liability." ECF No. 107-3 at 21.

26   Plaintiffs' arguments in response are unavailing. First, they cite to *ICONICS, Inc. v.*

27   *Massaro*, 192 F. Supp. 3d 254, 272 (D. Mass. 2016) for the proposition that "Section 1202(b) has

28   no requirement that the copy from which CMI is removed be identical." ECF No. 142 at 21. But

1   *ICONICS* concerned whether various "copyright headers" were CMI within the meaning of

2   Section 1202(c) because "the [plaintiff's] copyright at issue cover[ed] the full versions of

3   programs, not individual files." *ICONICS, Inc.*, 192 F. Supp. 3d at 272. Notably, the parties did

4   not dispute that the defendants reproduced an identical copy of the plaintiff's file. *Id.* Similarly,

5   Plaintiffs' attempt to liken this case to *Bounce Exchange, Inc. v. Zeus Enterprises Ltd*., No.

6   15CV3268, 2015 WL 8579023 (S.D.N.Y. Dec. 9, 2015) is unpersuasive. The question in *Bounce*

7   was whether two terms that appeared in the plaintiff's source code were CMI. *Id.* at *3. In

8   concluding that the terms were CMI, the *Bounce* court reasoned that the terms were "a shorthand

9   form of the official name of the author of the work, and they are inserted into the code itself[,]"

10   thereby satisfying the requirements of Section 1202(c). *Id.* In short, neither case cited by

11   Plaintiffs concerns Section 1202(b)'s identicality requirement.

12   Although the Court finds it unlikely that this deficiency could be cured by the allegation of

13   additional facts, it grants leave to amend out of abundance of caution. *See Rivas v. Kijakazi*, No.

14   C 23-03324 WHA, 2023 WL 8006846, at *2 (N.D. Cal. Nov. 17, 2023) (granting leave to amend

15   where "defects [in complaint] could theoretically be cured").

16   **CONCLUSION**

17   Defendants' motions to dismiss are denied in part and granted in part. Defendants'

18   motions to dismiss Plaintiffs' claims for damages for lack of standing under Article III are denied

19   as to Does 1, 2, and 5. Defendants' motions to dismiss Plaintiffs' claims for damages for lack of

20   standing under Article III are granted with prejudice as to Does 3 and 4. Moreover, Defendants'

21   motions to dismiss Plaintiffs' state law claims on preemption grounds, as well as their motions to

22   dismiss Plaintiffs' claims under Sections 1202(b)(1) and 1202(b)(3) of the DMCA, are granted.

23   Plaintiffs' state law claims for intentional and negligent interference with prospective economic

24   relations, unjust enrichment, negligence, and unfair competition are dismissed with prejudice.[9]

25   Plaintiffs' claims under Section 1202(b)(1) and 1202(b)(3) of the DMCA are dismissed with leave

26   to amend.

27

28

[9] Plaintiffs' unfair competition claim is dismissed only to the extent that it is predicated on their state law claims for intentional and negligent interference with prospective economic relations, unjust enrichment, and negligence.

16

1   Plaintiffs may file an amended complaint within 21 days of the date of this order. Leave to

2   amend is granted solely to correct the deficiencies identified in this order. Failure to file a timely

3   amended complaint will result in dismissal of all dismissed claims with prejudice.

4   **IT IS SO ORDERED.**

5   Dated: January 3, 2024

6   _____

7   JON S. TIGAR
                                                                    United States District Judge

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

17

EXHIBIT J

Joseph R. Saveri (State Bar No. 130064)
Cadio Zirpoli (State Bar No. 179108)
Christopher K.L. Young (State Bar No. 318371)
Louis A. Kessler (State Bar No. 243703)
Elissa A. Buchanan (State Bar No. 249996)
William W. Castillo Guardado (State Bar No. 294159)
Holden Benon (State Bar No. 325847)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1505
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
                czirpoli@saverilawfirm.com
                cyoung@saverilawfirm.com
                lkessler@saverilawfirm.com
                eabuchanan@saverilawfirm.com
                wcastillo@saverilawfirm.com
                hbenon@saverilawfirm.com

Counsel for Individual and Representative
Plaintiffs and the Proposed Class

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
OAKLAND DIVISION

J. DOE 1, et al.,
 	
Individual and Representative Plaintiffs,
 	
v.
 	
GITHUB, INC., et al.,
 	
                              Defendants.
	Case Nos. 4:22-cv-06823-JST
	                 4:22-cv-07074-JST
	
	PLAINTIFFS' NOTICE OF MOTION AND MOTION TO AMEND AND CERTIFY THE COURT'S JUNE 24, 2024 ORDER FOR INTERLOCUTORY APPEAL PURSUANT TO 28 U.S.C. § 1292(B); MEMORANDUM OF POINTS AND AUTHORITIES
	
	Hearing Date: September 26, 2024
	Time: 2:00 PM
	Judge: Hon. Jon S. Tigar
	Courtroom 6, 2nd Floor

TO THE COURT, PLAINTIFFS, AND THEIR COUNSEL OF RECORD:

PLEASE TAKE NOTICE THAT on September 26, 2024, at 2:00 PM. in Courtroom 6, 2nd Floor of the United States District Court for the Northern District of California, located at 1301 Clay Street, Oakland, California 94612, Plaintiff will move for an Order to amend and certify this Court's June 24, 2024 Order Granting in Part and Denying in Part Motions to Dismiss (ECF No. 253) for interlocutory appeal pursuant to 28 U.S.C. § 1292(b). This motion is based upon this Notice of Motion, the accompanying Memorandum of Points and Authorities, all other pleadings on file, and on such further written or oral argument as permitted by this Court.

Dated: July 24, 2024

By: */s/ Joseph R. Saveri*
                Joseph R. Saveri

Joseph R. Saveri (State Bar No. 130064)
Cadio Zirpoli (State Bar No. 179108)
Christopher K.L. Young (State Bar No. 318371)
Louis A. Kessler (State Bar No. 243703)
Elissa A. Buchanan (State Bar No. 249996)
William W. Castillo Guardado (State Bar No. 294159)
Holden Benon (State Bar No. 325847)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1505
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
                     czirpoli@saverilawfirm.com
                     cyoung@saverilawfirm.com
                     lkessler@saverilawfirm.com
                     eabuchanan@saverilawfirm.com
                     wcastillo@saverilawfirm.com
                     hbenon@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
Telephone: (323) 968-2632
Facsimile: (415) 395-9940
Email: mb@butticklaw.com

Counsel for Individual and Representative
Plaintiffs and the Proposed Class

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Cases

MOTION TO AMEND AND CERTIFY THE COURT'S JUNE 24, 2024 ORDER FOR INTERLOCUTORY APPEAL
PURSUANT TO 28 U.S.C. § 1292(B)

Statutes

MOTION TO AMEND AND CERTIFY THE COURT'S JUNE 24, 2024 ORDER FOR INTERLOCUTORY APPEAL
PURSUANT TO 28 U.S.C. § 1292(B)

Other Authorities

I.   INTRODUCTION

Plaintiffs seek to certify for interlocutory appeal a question of pure statutory interpretation in a class action of tremendous public import—precisely the type of legal issue for which interlocutory appeals were designed. Plaintiffs' class action complaint brings claims under 17 U.S.C. § 1202(b) of the Digital Millenium Copyright Act ("DMCA"), which collectively concern the removal of Copyright Management Information ("CMI") during both the Large Language Model ("LLM") training process and in the output generated by the LLM. Plaintiffs' DMCA claims are the heart of their class case: damages under the DMCA would potentially be in the billions of dollars. Whether CMI removal in the LLM training process is an issue of tremendous import is incontrovertible: Since Plaintiffs filed this case in 2022, at least five new lawsuits[1] have been brought against some of the same Defendants in this case, alleging analogous DMCA violations.

In this Court's June 24, 2024 Order Granting in Part and Denying in Part Motions to Dismiss (ECF No. 253) ("Third MTD Order"), it dismissed Plaintiffs' § 1202(b) claims, holding that those claims require that copies be "identical" in order for liability to attach. But for the Court's adoption of an "identicality" element to Plaintiffs' §§ 1202(b)(1) and (b)(3) claims—a term and legal standard that appears nowhere in § 1202 or its legislative history—Plaintiffs' DMCA claims would still be in this case. The question of whether claims under § 1202(b)(1) or (b)(3) contain an element of "identicality" is a question ripe for interlocutory appeal.

It is readily apparent that the first and second elements for interlocutory appeal under 28 U.S.C. § 1292 are met. This is a controlling question of law as to which there is a substantial ground for difference of opinion upon which reasonable jurists can disagree. This is not merely a case where Plaintiffs vigorously disagree with the Court's ruling.

---

[1] *See Concord Music Group, Inc. v. Anthropic PBC*, in the U.S. District Court for the Middle District of Tennessee, Nashville Division, No. 3:23-cv-01092 (filed Oct. 18, 2023) (same); *The New York Times Co. v. Microsoft Corp.*, in the U.S. District Court for the Southern District of New York, 1:23-cv-11195 (filed Dec. 27, 2023) (alleging violations of § 1202(b)); *Huckabee v. Bloomberg*, in the U.S. District Court for the Southern District of New York, 1:23-cv-11195 (filed Dec. 27, 2023) (same); *Raw Story Media, Inc. v. OpenAI*, in the U.S. District Court for the Southern District of New York, 1:24-cv-01514 (filed Feb. 28, 2024) (alleging violations of § 1202(b)); *Intercept Media Inc. v. OpenAI*, in the U.S. District Court for the Southern District of New York, 1:24-cv-01515 (filed Feb. 28, 2024) (same).

1   As a starting point, there is no dispute that the plain language of § 1202 itself does not require that

2   a copy be "identical." The only instance of the word "identical" is found in another provision of the

3   DMCA, 17 U.S.C. § 1201(d)(2), which provides for an exemption for liability for certain public institutions

4   under that section of the statute. That Congress chose to include the word "identical" in one section of

5   the DMCA but not the other presumes Congress acted intentionally and purposely in the disparate

6   inclusion or exclusion. Indeed, nothing in the legislative materials leading up to the DMCA's passage into

7   law suggests Congress had a different intent.

8   Setting aside the plain language of the statute, there is no controlling case law on this point. No

9   Court of Appeals has spoken on this issue. The Court's own analysis of contrary precedent confirms that

10   "there is substantial ground for difference of opinion," as to whether §§ 1202(b)(1) and (b)(3) includes an

11   "identicality" element. Though the Ninth Circuit has not spoken on this point specifically, it has implicitly

12   rejected an "identicality" standard for the DMCA. Given the lack of clarity from the federal circuit courts,

13   district courts have been sharply divided. The lack of clarity on this critical question for § 1202 claims are

14   precisely the questions ripe for interlocutory appeal.

15   The third element for review under § 1292(b) is also met: resolution by the Ninth Circuit will

16   materially advance the ultimate termination of the litigation. Importantly, this is a class case wherein

17   Plaintiffs' key claim has been dismissed. No one disputes that the bulk of Plaintiffs' damages lie in their

18   DMCA claim. And early resolution of that key claim will advance the final resolution of this litigation. As

19   explained by Judge Posner, "[t]hat is enough to satisfy the 'may materially advance' cause of section

20   1292(b)[.]" Further, appeal will result in little delay. Although pending for almost two years, this case is

21   still in its infancy with respect to its procedural posture. Discovery has not meaningfully advanced, given

22   the centrality of Plaintiffs' DMCA claims. Indeed, due to those claims' importance, even were Plaintiffs

23   to prevail at trial, a separate trial would likely be required again should this Court's decision be reversed.

24   But there is more. The certification of these "controlling questions of law" to the Ninth Circuit stretches

25   well beyond the instant litigation. Guidance from the Ninth Circuit will not only settle the law of this

26   Circuit but will also materially affect several analogous cases pending both within and outside this Circuit.

27   For these reasons and others set forth below, Plaintiffs have met their burden of demonstrating that

28   the Court's Third MTD Order meets all three elements for § 1292(b) certification and presents the kind

of extraordinary circumstance that warrants interlocutory review. Plaintiffs therefore respectfully seek certification of this Court's June 24, 2024 Order for interlocutory appeal pursuant to 28 U.S.C. § 1292(b), and a stay pending appeal.

II.   RELEVANT BACKGROUND

On November 21, 2022, Plaintiffs filed their initial complaint in this matter alleging, as relevant here, that Defendants violated § 1202(b)(1) and §1202(b)(3) by removing or altering Copyright Management Information ("CMI") from Plaintiffs' licensed software code and distributing copies of that licensed code without including its CMI. ECF No. 1 ¶¶ 138–167. On May 11, 2023, the Court issued its First MTD Order upholding Plaintiffs' DMCA claims under §1202(b). ECF No.95 ("First MTD Order").

On June 8, 2023, Plaintiffs filed their First Amended Complaint (ECF No. 97) ("FAC"), which included new facts to address deficiencies the Court identified in other causes of action First Amended Complaint, ECF No. 97. When moving to dismiss the FAC, Defendants asked the Court to reconsider its decision with respect to Plaintiffs' previously upheld §1202(b) claims, arguing that such claims failed to allege facts sufficient to show that Defendants' copying was "identical" under the DMCA. ECF Nos. 107-2, 109-3. In this Court's January 3, 2024 Order Granting in Part Denying in Part Motion to Dismiss (ECF No. 189) ("Second MTD Order"), the Court found that the new facts Plaintiffs alleged showed that Plaintiffs had failed to plead that the Defendants had reproduced Plaintiffs' work in violation of the DMCA. In particular, the Court held that the DMCA required Plaintiffs to plead facts sufficient to show that output from Copilot is an "identical copy." ECF No. 189 at 15. The Court noted that Plaintiffs did not allege that Defendants had produced identical copies of Plaintiffs' works. *Id.* at 14-15. Though the Court recognized that § 1202(b)(1) only requires intentionally removing or altering CMI without the authority of the copyright owner, and is therefore distinct from § 1202(b)(3) (which makes it illegal to "distribute, [or] import for distribution . . .. . . copies of works" with CMI removed or altered), the Court attached an "identicality" requirement to claims under both DMCA §§ 1202(b)(1) and (b)(3). ECF No. 189 at 15. Plaintiffs were given leave to amend to meet this standard.

On January 25, 2024, Plaintiffs filed their Second Amended Complaint (EFC No. 200) ("SAC"), which included, among other facts, new allegations showing the likelihood that Plaintiffs' and class members' licensed code would be emitted verbatim overtime as the capacity and scope of Copilot grew.

ECF No. 200 at ¶¶ 104-07. On June 24, 2024, this Court reconfirmed its dismissal of Plaintiffs' §1202(b)

claims because Plaintiffs had "failed to meet the DMCA's identicality requirement." ECF No. 253 at 5.

This Court ruled that *ADR Int'l Ltd. v. Inst. for Supply Mgmt. Inc.,* 667 F. Supp. 3d 411 (S.D. Tex. 2023)("

ADR Int'l") was not binding, and instead relied on decisions within the Ninth Circuit to support the

conclusion that §1202(b) included an "identicality" requirement. *Id.* at 4–5. As Plaintiffs argued, those

cases did not support such a conclusion because they either never addressed the issue of identicality, relied

on faulty reasoning, or misapplied or misinterpreted relevant case law.

III.   ARGUMENT

A "district court should not hesitate to certify an interlocutory appeal" when its ruling "[i]nvolves

a new legal question or is of special consequence" *Mohawk Indus., Inc. v. Carpenter*, 558 U.S. 100, 110-11

(2009). "The trial judge has discretion to certify a decision for interlocutory review if all of the following

statutory elements are met: (1) the order 'involves a controlling question of law'; (2) there is 'substantial

ground for difference of opinion'; and (3) 'an immediate appeal from the order may materially advance the

termination of the litigation.'" *United States ex rel. Integra Med. Analytics LLC v. Providence Health & Servs.*,

No. 17-1694, 2019 WL 6973547, at *4 (C.D. Cal. Oct. 8, 2019) (quoting 28 U.S.C. § 1292(b)). The Court's

Third MTD Order granting Defendants' motion to dismiss Plaintiffs' DMCA claims satisfies each of these

elements such that there are compelling reasons to depart from the general rule disfavoring piecemeal

appeals. Specifically, the Court's Third MTD Order includes the following two "controlling question[s]

of law," 28 U.S.C. § 1292(b):

1. Whether an "identicality" standard applies to Plaintiff's §1202(b)(1) claim, which prohibits the "intentional[] remov[al] or alter[ation of] any copyright management information" from a work while "knowing, or . . . having reasonable grounds to know, that it will induce, enable, facilitate, or conceal an infringement of any right under this title"; and

2. Whether an "identicality" standard applies to Plaintiffs' §1202(b)(3) claim, which prohibits the "distribut[ion] or import for distribution . . . [of] copies of works . . . knowing that copyright management information has been removed or altered" while "knowing, or . . . having reasonable grounds to know, that it will induce, enable, facilitate, or conceal an infringement of any right under this title."

For the reasons discussed below, this Court should certify for interlocutory appeal its dismissal of

Plaintiffs' §1202(b) claims.

A. Whether Sections 1202(b)(1) or (b)(3) Include an "Identicality" Requirement is a Controlling Question of Law

"[A]ll that must be shown in order for a question to be 'controlling' is that resolution of the issue on appeal could materially affect the outcome of litigation in the district court." *In re Cement Antitrust Litig.*, 673 F.2d 1020, 1026 (9th Cir. 1982). The "controlling question of law in an interlocutory appeal generally is a purely legal one that can be resolved quickly without delving into a particular case's facts," *Henley v. Jacobs*, No. C 18-2244 SBA, 2019 WL 8333448, at *2 (N.D. Cal. Oct. 25, 2019) (citing *Steering Comm. v. United States*, 6 F.3d 572, 575-76 (9th Cir. 1993)), and "need not be dispositive of a litigation," *See Adams v. Cnty. of Sacramento*, No. 222CV01499WBSKJN, 2023 WL 3413672, at *1 (E.D. Cal. May 12, 2023). Such "purely legal" questions are appropriate for interlocutory appeal because they involve abstract legal issues which the court of appeals can resolve quickly and cleanly. The best examples of this are the meaning of a statute, regulation, or constitutional provisions. *See, e.g.*, *Boniface v. Viliena*, 417 F. Supp. 3d 113, 123 (D. Mass. 2019) ("A controlling question of law usually involves a question of the meaning of a statutory or constitutional provision, regulation, or common law doctrine rather than an application of law to the facts." (citations omitted)). "In the absence of controlling Ninth Circuit precedent, and in the presence of conflicting authority elsewhere, the Ninth Circuit should have the opportunity to review this question of law." *Synthesis Indus. Holdings I, LLC v. U.S. Bank Nat'l Ass'n*, Case no. 2-19-CV-1431 JCM, 2021 WL 2406895, at *2 (D. Nev. June 11, 2021) (citation and internal quotation marks omitted).

The plain language of § 1202(b) does not include the word "identical." Indeed, the DMCA contains only one mention of the word "identical"—as an exemption under Section 1201 for nonprofit libraries, archives, and educational institutions. *See* 17 U.S.C. § 1201(d)(2) ("The exemption made available under paragraph (1) shall only apply with respect to a work when an **identical** copy of that work is not reasonably available in another form.") (emphasis added). The inclusion of the word "identical" in § 1202's sister provision of the DMCA compels a singular conclusion—if Congress wanted to include an "identicality" element for § 1202 claims, it would have. *Keene Corp. v. United States*, 508 U.S. 200, 208 (1993) ("[W]here Congress includes particular language in one section of a statute but omits it in another . . . , it is generally presumed that Congress acts intentionally and purposely in the disparate inclusion or exclusion.") (alteration in original, internal quotation marks omitted, citing *Russello v. United States*, 646

U.S. 16, 23 (1983)); *see also Pettis ex rel. U.S. v. Morrison-Knudsen Co., Inc.*, 577 F.2d 668 672 (9th Cir. 1978) ("We have no doubt but that under such circumstances the intent of Congress resides in the words of the statute. That is, discharge of our obligation to follow the intent of Congress requires that we assume that Congress said what it meant and meant what it said."). The legislative history of the DMCA is in accord. Indeed, nothing in the legislative materials leading up to the passage of the DMCA suggests any other intent by Congress. S. Rep. 105-190, at 31 (1998). Nonetheless, this Court reached a different conclusion.

Whether §§ 1202(b)(1) or (b)(3) requires parties to plead and prove "identicality" between the original and the copy is a purely legal question as it concerns a matter of statutory interpretation. *In re Google Inc. St. View Elec. Commc'ns Litig.*, No. C 10-MD-02184 JW, 2011 WL 13257346, at *1 (N.D. Cal. July 18, 2011) (finding a "controlling question of law" suitable for interlocutory appeal regarding a "novel question of statutory interpretation"); *San Antonio Winery, Inc. v. Jiaxing Micarose Trade Co.*, No. CV 20-9663-GW-KSX, 2021 WL 4988033, at *3 (C.D. Cal. June 1, 2021) ("The Court agrees that the language of 15 U.S.C. § 1051(e) is ambiguous and subject to statutory interpretation which is strictly a legal question. And because the statute is subject to interpretation, courts have made opposing decisions."); *Synthesis Indus. Holdings*, 2021 WL 2406895, at *2 (finding a "controlling question of law" where it involved "one of pure statutory interpretation, the interpretation and application of Fed. R. Bankr. P. 7004(h)."). Moreover, the Court of Appeals can resolve this question quickly and cleanly without delving into the factual record, and resolution of this purely legal question in favor of Plaintiffs will materially impact the outcome of litigation by reviving a claim Plaintiffs will otherwise appeal after a final judgement.

B. Differences of Opinion as to Section 1202(b)'s "Identicality" Requirement Already Exist and are Likely to Proliferate

Also satisfied is §1292(b)'s requirement for "substantial ground for difference of opinion" as to the meaning of the statute. "Courts traditionally will find that a substantial ground for difference of opinion exists where 'the circuits are in dispute on the question and the court of appeals of the circuit has not spoken on the point, if complicated questions arise under foreign law, or if novel and difficult questions of first impression are presented.'" *Silbersher v. Allergan Inc.*, No. 18-CV-03018-JCS, 2021 WL 292244, at *2 (N.D. Cal. Jan. 28, 2021) (quoting *Couch v. Telescope Inc.*, 611 F.3d 629, 633 (9th Cir. 2010) (citation omitted)). Notably, the Ninth Circuit has stated that "when novel legal issues are presented, on which fair-

minded jurists *might* reach contradictory conclusions, a novel issue may be certified for interlocutory appeal without first awaiting development of contradictory precedent." *Reese v. BP Expl. (Alaska) Inc.*, 643 F.3d 681, 688 (9th Cir. 2011) (italics added). "[I]dentification of a sufficient number of conflicting and contradictory opinions would provide substantial ground of disagreement." *Ray v. Cal. Dep't of Soc. Servs.*, Case No. CV 17-4239 PA (SKx), 2017 WL 10436062, at *3 (C.D. Cal. Nov. 20, 2017) (quoting *Union Cnty., Iowa v. Piper Jaffray & Co., Inc.*, 525 F.3d 643 (8th Cir. 2008), cited with approval in *Couch*, 611 F.3d at 633-34); *see also Rollins v. Dignity Health*, 2014 WL 6693891, at *3 (N.D. Cal. Nov. 26, 2014) ("One of the best indications that there are substantial grounds for disagreement on a question of law is that other courts have, in fact, disagreed.") (citing *Couch*, 611 F.3d 629, 633 and *Reese*, 653 F.3d at 688).

Whether §§ 1202(b)(1) or (b)(3) include an "identicality" requirement is a novel issue on which no Court of Appeals has directly spoken. Notably, the Ninth Circuit has implicitly rejected an "identicality" standard for the DMCA. *See Friedman v. Live Nation Merch., Inc.*, 833 F.3d 1180, 1188 (9th Cir. 2016) (holding a "'striking similarity' between the works may give rise to a permissible inference of copying" supporting a DMCA claim). Further, district courts are sharply divided.

This Court chiefly relied on *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-1931-DMG (EX), 2020 WL 5991503, at *6 (C.D. Cal. Aug. 14, 2020) in support of its conclusion that an "identicality" requirement adheres to §§1202(b)(1) and (b)(3) claims. But as noted in *ADR Int'l*, "[a]lthough the court in *Kirk Kara* held the DMCA requires identical copies, the case law it cited does not support its holding." *ADR Int'l*, 667 F. Supp. 3d at 427. For example, the *Kirk Kara* court pointed to *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999) and *Frost-Tsuji Architects v. Highway Inn, Inc.*, No. 13-00496, 2015 WL 263556, at *2 (D. Haw. Jan. 21, 2015), but neither case mentioned nor employed an identical copies requirement under the DMCA. The *Kirk Kara* court similarly cited *Fischer v. Forrest*, 286 F. Supp. 3d 590 (S.D.N.Y. 2018); yet the term "identical" appears nowhere in that decision. *Fischer* also did not hold identicality was required—in the *Fischer* court's brief discussion of the issue, it indicated that Plaintiffs had not plead that "the underlying work ha[d] been *substantially* or entirely reproduced." *Id.* at 609 (emphasis added); *see also ADR Int'l*, 667 F. Supp. 3d at 427 (rejecting *Kirk Kara*'s reliance on *Fischer*).

District courts, including those within the Ninth Circuit, have held DMCA liability can attach even when the work at issue itself is not an "exact copy" or even when the copy is "altered," implying the copied

work is not identical. *See, e.g.*, *Oracle Int'l Corp. v. Rimini St., Inc.*, No. 214CV01699MMDDJA, 2023 WL 4706127, at *82 (D. Nev. July 24, 2023) (rejecting argument "that a work that removes copyright management information must be an *exact copy* of the original work") (emphasis added); *see also, e.g.*, *Software Pricing Partners, LLC v. Geisman*, No. 319CV00195RJCDCK, 2022 WL 3971292, at *5 (W.D.N.C. Aug. 31, 2022) ("Here, Geisman *altered* SPP's copyrighted documents by removing indications of the copyright or otherwise altering the documents prior to distributing to customers. Geisman, as a former employee of SPP, reasonably knew that such information was copyrighted work and knew he was altering it by changing it enough to look like his own work while maintaining a substantial similarity to the original work. Accordingly, Geisman is liable for violating the DMCA.") (emphasis added, citations omitted).

The Court's adoption of an "identicality" element for § 1202 claims is also in tension with other cases addressing DMCA violations with respect to software code specifically. In those cases, courts have recognized that CMI can be embedded within computer code itself, so removal of that CMI would necessarily mean that the infringing copy would no longer be identical. *Oracle Int'l*, 2023 WL 4706127, at *82 ("Courts have held that when a defendant 'modifie[s] source code "substantially similar" to Plaintiff's copyrighted source code,' including by replacing the author's name with its own, the defendant is liable under the DMCA" citation omitted)); *Bounce Exch., Inc. v. Zeus Enter., Ltd.*, No. 15cv3268 (DLC), 2015 WL 8579023, at *3 (S.D.N.Y. Dec. 9, 2015) (finding § 1202(b) liability for removing CMI that was woven into and incorporated into code). At least one court in this circuit has determined that a § 1202 claim was adequately pleaded when the copy of the software at issue was a "derivative"—and derivatives are, not identical. *Splunk, Inc. v. Cribl, Inc.*, 662 F. Supp 3d 1029, 1053-54 (N.D. Cal. 2023).

The foregoing indicates that reasonable jurists might not reach the same conclusion as this Court. The divided conclusions of district courts, the lack of guidance from any circuit (let alone the Ninth Circuit), and the contrary conclusions reached by the *ADR Int'l* and *Fischer* courts (among others), confirms that "substantial grounds for differences of opinion" on the correct legal standard already exist. *See, e.g.*, *Rollins*, 2014 WL 6693891, at *3 (finding "substantial grounds for disagreement" where "two district courts have decided this issue explicitly in conflict with this Court's decision"); *Camacho v. Bridgeport Fin. Inc.*, 2004 WL 7336833, at *1 (N.D. Cal. July 16, 2004) (finding "there is a substantial ground for difference of opinion as to a controlling question of law" where "[d]istrict courts have split

widely on the matter" and where "[t]here is no controlling authority in the Ninth Circuit").

C. An Appeal will Materially Advance this Litigation and Others Like It

"[N]either § 1292(b)'s literal text nor controlling precedent requires that the interlocutory appeal have a final, dispositive effect on the litigation, only that it 'may materially advance' the litigation." *Reese*, 643 F.3d at 988 (citing 28 U.S.C. § 1292(b)). The Ninth Circuit settling this legal standard would materially advance not only this particular class action litigation, but numerous others challenging LLM models under the DMCA across the country. *J. B. v. G6 Hosp.*, No. 19-CV-07848-HSG, 2021 WL 6621068, at *4 (N.D. Cal. Dec. 16, 2021) ("Rather than litigating the case to the finish under a standard that will be challenged on appeal, the Court and the parties will benefit from definitive guidance from the Ninth Circuit at the outset, before time and resources are invested."). Indeed, "[w]hether an appeal may materially advance the termination of the litigation is 'linked to whether an issue of law is "controlling" in that the court should consider the effect of a reversal by the Ninth Circuit on the management of the case.'" *In re Cal. Title Ins. Antitrust Litig.*, No. 08-01341 JSW, 2010 WL 785798, at *2 (N.D. Cal. Mar. 3, 2010). And certification of interlocutory appeal will materially advance the litigation for at least four reasons:

First, courts have certified questions for interlocutory appeals where the "issues go to the heart of the case." *United States ex rel Huangyan Import & Export Corp. v. Nature's Farm Prods., Inc.*, 370 F. Supp. 2d 993, 1005 (N.D. Cal. 2005). Plaintiffs' § 1202 claims (their remaining federal claims) are certainly at the core of this case. *See Huangyan Import & Export Corp.*, 370 F. Supp. 2d at 1005 ("Depending on how the three issues are resolved [on interlocutory appeal], the United States might have two FCA claims, one FCA claim, no claim at all or might be in the wrong court altogether."). And that is before considering the stakes—should Plaintiffs prevail on their DMCA claim, they would be entitled to recover statutory damages "in the sum of not less than $2,500 or more than $25,000" per violation. 17 U.S.C. § 1206(c)(3)(B). And Plaintiffs have alleged tens of thousands, if not more, individual violations, amounting to potentially billions in damages. *Huangyun Import & Export Corp.*, 370 F. Supp. 2d at 1005 ("Furthermore, the stakes are large—after trebling and civil penalties, there are tens of millions of dollars in controversy."); *Casas v. Victoria's Secret Stores, LLC*, 2015 WL 13446989, at *3 (C.D. Cal. Apr. 9, 2015) (certifying interlocutory appeal and noting that an estimated $37 million amount-in-controversy for claim at issue "is not a 'life-jacket' claim; it is central to Plaintiffs' case."); *S.E.C. v. Mercury Interactive, LLC*, Case No. 5:07-

cv-02822-JF, 2011 WL 1335733, at *3 (N.D. Cal. Apr. 7, 2011) (certifying appeal where "[t]he bulk of the damages sought against Defendants arise from the § 304 claims"); *see also Sterk v. Redbox Automated Retail, LLC*, 672 F.3d 535, 536 (7th Cir. 2012) ("[T]he completion of the litigation will take longer than if the destruction claim is out of the case, especially since that claim appears to be the plaintiffs' main one, with the disclosure claim perhaps just a life jacket.") (Posner, J.).

Second, and relatedly, this is a class action. The grant of an interlocutory appeal would "advance termination of litigation by increasing odds of settlement," an efficiency gain courts have recognized is particularly relevant in the context of "class action cases, where litigants would be able to more accurately predict their odds of success, scope of potential liability, and a fair estimate of the case's value." *See In re N. Dist. of Cal., Dalkon Shield IUD Prods. Liab. Litig.*, 526 F. Supp 887, 919 (N.D. Cal. 1981) (recognizing that an appeal may materially advance termination where "immediate resolution . . . will lead to increased settlements and save thousands of hours of court time"). As observed by one court in the Ninth Circuit, "especially in class actions, uncertainty over a key claim's status 'may delay settlement (almost all class actions are settled . . .), and by doing so further protract the [case].'" *Casas*, 2015 WL 13446989, at *3 (quoting *Sterk*, 672 F.3d at 536 (Posner, J.)). "As Judge Posner explained in *Sterk*, '[t]hat is enough to satisfy the "may materially advance" cause of section 1292(b)[.]'" *Id.*; *see also Canela v. Costco Wholesale Corp.*, Case No. 13-cv-03598-BLF, 2018 WL 3008532, at *2 (N.D. Cal. June 15, 2018) ("If Canela is limited to pursuing only her individual PAGA claim . . . , the trial would involve fewer disputed issues and it would be more likely that the parties would reach a settlement given Costco's willingness to settle this case under that circumstance."); *Rollins*, 2014 WL 6693891, at *4 ("By addressing the questions now, the Court saves time and expense. If the Ninth Circuit reverses, the parties can turn to these issues sooner rather than later. And if the Court of Appeals affirms, the case can proceed on the relatively few issues that remain with greater certainty. Such certainty could even encourage a negotiated settlement, which would not just materially but completely advance the termination of this litigation.").

Third, courts have routinely found this requisite met where the grant of an appeal "would resolve a legal issue implicated in a large number of other cases." *See, e.g.*, *Leite v. Crane Co.*, No. 11–00636 JMS/RLP, 2012 WL 1982535, at *7 (D. Haw. May 31, 2012) (collecting cases recognizing how the impact of an interlocutory appeal on other cases is a factor that courts "may take into account in deciding whether

to accept an appeal that has been properly certified by the district court"); *Krangel v. Crown*, 791 F.Supp. 1436, 1449 (S.D. Cal. 1992) ("Certification for appeal may also materially advance the conclusion of other cases involving this same legal issue."); *In re Methyl Tertiary Butyl Ether Prods. Liability Litig.*, 399 F. Supp. 2d 320, 324 (S.D.N.Y.2005) (stating that courts consider, in part, whether the certified issue has precedential value for a large number of cases); *Klinghoffer v. S.N.C. Achille Lauro*, 921 F.2d 21, 24 (2d Cir.1990) ("[T]he impact that an appeal will have on other cases is a factor that we may take into account in deciding whether to accept an appeal that has been properly certified by the district court."). Numerous other cases challenging other LLMs, almost all of them class actions, have brought similar DMCA claims. *See* note 1, *supra*. Those similarly situated will benefit from the resolution of these issues on interlocutory appeal. *See Ass'n of Irritated Residents v. Fred Schakel Dairy*, 634 F. Supp. 2d 1081, 1093 (E.D. Cal. 2008) ("The opportunity to achieve appellate resolution of an issue important to other similarly situated dairies can provide an additional reason for certification . . . .") (citing *Klinghoffer*, 921 F.2d at 24 (2d Cir. 199); 16 *Wright, Miller & Coper*, Fed. Prac. & Proc.*, § 3930, p. 425).

       Fourth, this case is still in an early procedural stage. Given the import of Plaintiffs' § 1202 claims, resolution of this question would eliminate duplication should the Court's dismissal of Plaintiffs' § 1202 claims be reversed after trial. *Casas*, 2015 WL 13446989, at *4("Given the centrality of Plaintiffs' call-in reporting-time claim to their overall case, and the early stage at which the issue presents itself, the Court would conclude that Plaintiffs have established that an appeal 'may materially advance' the ultimate resolution without unduly delaying the currently-planned course of litigation."). At this juncture, discovery has not materially advanced. Little effort to date would need to be duplicated and, indeed, in the event of reversal, much duplicative work to come would be prevented. *Adams v. Cnty. of Sacramento*, No. 2:22-cv-01499 WBS KJN, 2023 WL 3413672, at *2 (E.D. Cal. May 12, 2023) (stating that a "final decision on a controlling legal issue sooner, rather than later [will] save the courts and litigants unnecessary trouble and expense"). But if this case were to proceed, after which this Court's order is reversed, the parties and the Court's efforts would be undoubtedly duplicated. *See Sterk*, 672 F.3d at 536; *see also Adams*, 2023 WL 3413672, at *2 ("[A]n immediate appeal could avoid the need for two separate trials in the event this court's dismissal of Claims 3 and 4 is reversed."); *Mercury Interactive*, 2011 WL 1335733, at *3 ("A final resolution as the scope of the statute would have a significant effect on the trial of this action, and perhaps upon the

MOTION TO AMEND AND CERTIFY THE COURT'S JUNE 24, 2024 ORDER FOR INTERLOCUTORY APPEAL PURSUANT TO 28 U.S.C. § 1292(B)

parties' efforts to reach settlement."); *Thompson v. Procter & Gamble Co.*, Case No. C-80-3711 EFL, 1982 WL 114, at *2 (N.D. Cal. Dec. 8, 1982) ("The primary reason that this Court requests review is the potential waste of judicial time and effort which would result from reversal after trial.").

Give the foregoing, the Court should certify for interlocutory appeal its Third MTD Order so that the Ninth Circuit can clarify whether §§1202(b)(1) and (b)(3) includes an identicality standard.

D. The District Court Should Enter of Stay of Proceedings Pending Appeal

"A district court 'has broad discretion to stay proceedings as an incident to its power to control its own docket' in an effort to promote judicial economy." *DeMartini v. Johns*, 693 F. App'x 534, 538 (9th Cir. 2017) (quoting *Clinton v. Jones*, 520 U.S. 681, 706-07 (1997)). District courts may also order a stay of proceedings pending an interlocutory appeal. *See* 28 U.S.C. § 1292(b). Indeed, "a district judge presiding over an action from which interlocutory appeal has been granted may exercise its discretion to impose a stay of proceedings if such a stay would 'promote economy of time and effort for itself, for counsel, and for litigants." *Finder v. Leprino Foods Co.*, No. 13-cv-2059, 2017 WL 1355104, at *1 (E.D. Cal. Jan. 20, 2017) (quoting *Filtrol Corp. v. Kelleher*, 467 F.2d 242, 244 (9th Cir. 1972)).

This Court has held that the test derived from *Landis v. North American Co.*, 299 U.S. 248 (1936) provides the appropriate standard under which to consider a request to stay proceedings pending an interlocutory appeal. *Kuang v. U.S. Dep't of Defense*, No. 18-cv-3698, 2019 WL 1597495, at *2-4 (N.D. Cal. Apr. 15, 2019) (Tigar, J.). "Under this test, courts examine (1) 'the possible damage which may result from the granting of a stay'; (2) 'the hardship or inequity which a party may suffer [if the case is allowed] to go forward'; and (3) 'the orderly course of justice measured in terms of the simplifying or complicating of issues, proof, and questions of law which could be expected to result from a stay.'" *Id.* at *2 (alteration in original). Application of all three factors weighs in favor of granting a stay of proceedings pending appeal.

With respect to the first two *Landis* factors, no damage will result to Defendants if the Court enters a stay. The parties are still in the early stages of discovery: less than one-thousand documents have been produced to date by the three Defendants combined. No depositions have been taken. A stay of the proceedings would merely pause a case that has been moving slowly through the motion to dismiss process. In contrast to the lack of harm to Defendants, however, failing to grant a stay will cause great hardship and inequity to Plaintiffs. If the Ninth Circuit agrees with Plaintiffs that §1202(b) does not include an

"identicality" standard, Plaintiffs' DCMA claims will be revived given that this Court upheld them initially. Plaintiffs will then be forced to redo discovery, expert work, and class certification, which will be costly and burdensome. *See, e.g.*, *Gustavson v. Mars, Inc.*, No. 13-CV-04537-LHK, 2014 WL 6986421, at *3 (N.D. Cal. Dec. 10, 2014) (finding these factors met because "the need to re-brief class certification and potentially re-open discovery would involve a significant expenditure of time and resources").

The third *Landis* factor similarly favors a stay of the proceedings. Were the Ninth Circuit to disagree with this Court's interpretation of the appropriate legal standard for §§1202(b)(1) and (b)(3) claims, considerable judicial resources will be expended as the parties are forced to relitigate claims on a bifurcated timeline. "It would be a waste of judicial and party resources to proceed with the other claims while the appeal is pending." *Ass'n of Irritated Residents*, 634 F. Supp. 2d at 1094.

IV.   CONCLUSION

For the foregoing reasons, the Court should certify for interlocutory appeals the questions of whether claims brought under §§ 1202(b)(2) and (b)(3) require an element of "identicality." The Court should also enter a stay pending the resolution of the interlocutory appeal.

Dated: July 24, 2024   By: */s/ Joseph R. Saveri*
                                                   Joseph R. Saveri

                                        Joseph R. Saveri (State Bar No. 130064)
                                        Cadio Zirpoli (State Bar No. 179108)
                                        Christopher K.L. Young (State Bar No. 318371)
                                        Louis A. Kessler (State Bar No. 243703)
                                        Elissa A. Buchanan (State Bar No. 249996)
                                        William W. Castillo Guardado (State Bar No. 294159)
                                        Holden Benon (State Bar No. 325847)
                                        JOSEPH SAVERI LAW FIRM, LLP
                                        601 California Street, Suite 1505
                                        San Francisco, California 94108
                                        Telephone: (415) 500-6800
                                        Facsimile: (415) 395-9940
                                        Email: jsaveri@saverilawfirm.com
                                                       czirpoli@saverilawfirm.com
                                                       cyoung@saverilawfirm.com
                                                       lkessler@saverilawfirm.com
                                                       eabuchanan@saverilawfirm.com
                                                       wcastillo@saverilawfirm.com
                                                       hbenon@saverilawfirm.com

                                        Matthew Butterick (State Bar No. 250953)
                                        1920 Hillhurst Avenue, #406
                                        Los Angeles, CA 90027
                                        Telephone: (323) 968-2632
                                        Facsimile: (415) 395-9940
                                        Email: mb@butoricklaw.com

                                        *Counsel for Individual and Representative
                                        Plaintiffs and the Proposed Class*