ANNETTE L. HURST (SBN 148738)
ahurst@orrick.com
ORRICK, HERRINGTON & SUTCLIFFE LLP
405 Howard Street
San Francisco, CA 94105-2669
Telephone: +1 415 773 5700
Facsimile: +1 415 773 5759

WILLIAM W. OXLEY (SBN 136793)
woxley@orrick.com
ALYSSA CARIDIS (SBN 260103)
acaridis@orrick.com
ORRICK, HERRINGTON & SUTCLIFFE LLP
355 S. Grand Avenue
Los Angeles, CA 90071
Telephone: +1 213 629 2020
Facsimile: +1 213 612 2499

Attorneys for GitHub, Inc. and Microsoft Corporation

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.

AND CONSOLIDATED ACTION

Case No. 4:22-cv-6823-JST

Consolidated with Case No. 4:22-cv-7074-JST

DEFENDANTS GITHUB AND MICROSOFT'S NOTICE OF MOTIONS AND MOTIONS TO DISMISS OPERATIVE COMPLAINT IN CONSOLIDATED ACTIONS

Date: May 4, 2023
Time: 2:00 p.m.
Courtroom: 6, 2d Floor
Judge: Hon. Jon S. Tigar

Complaint Filed: December 7, 2022

TABLE OF CONTENTS

TABLE OF AUTHORITIES iii

NOTICE OF MOTIONS AND MOTIONS viii

MEMORANDUM OF POINTS AND AUTHORITIES 1

INTRODUCTION AND SUMMARY OF ISSUES 1

ALLEGATIONS OF THE OPERATIVE COMPLAINT 2

A. OpenAI Develops A Generative AI Tool Called Codex 2

B. GitHub Offers Copilot, A Code Completion Tool Built On Codex. 3

C. Plaintiffs Sue Based On An Attribution Theory. 4

ARGUMENT 5

I. PLAINTIFFS LACK ARTICLE III STANDING AND THEREFORE SUBJECT MATTER JURISDICTION BECAUSE THEY HAVE NOT ALLEGED ACTUAL OR THREATENED INJURY. 5

- A. Plaintiffs' Lack-Of-Attribution Theory Is Insufficient To Confer Standing. 6
- B. Plaintiffs Also Do Not Allege Privacy-Based Injury. 8

II. PLAINTIFFS FAIL TO STATE A DMCA CLAIM. 8

- A. Plaintiffs Cannot Allege The Required Likelihood Of Infringement. 9
- B. Count I Impermissibly Lumps Together Multiple Defendants, Claims, And Theories Of Liability. 12
- C. Plaintiffs Have Not Plausibly Alleged That GitHub Or Microsoft Intentionally Or Knowingly Removes Or Alters CMI Under § 1202(b). 13
- D. Plaintiffs Have Not Plausibly Alleged That GitHub Or Microsoft Provide, Distribute, Or Import For Distribution Any False CMI. 14

III. PLAINTIFFS FAIL TO STATE A BREACH-OF-LICENSE CLAIM. 15

- A. Plaintiffs Fail To Specify The Contract Provisions Allegedly Breached. 16
- B. To The Extent Count II Is Based On Training Copilot With Code Found In Public Repositories, It Is Foreclosed By GitHub's Terms Of Service. 16
- C. Plaintiffs Have Not Plausibly Alleged Breach Based On An Output Theory 17

IV. PLAINTIFFS' TORT AND UCL CLAIMS FAIL. 18

- A. Plaintiffs' Tort And UCL Claims Are Preempted By The Copyright Act. 18

B. Plaintiffs Do Not Plausibly Allege A Contract Or Business Expectancy, Nor Defendants' Knowledge Of One As Required For An Interference With Economic Advantage Claim. 20

C. Plaintiffs' Fraud Claim Against GitHub Is Barred By The Economic Loss Rule And Fails Rule 9. 21

D. Plaintiffs Lack Standing To Bring A UCL Unlawful Competition Claim And Fail To State A Claim. 22

V. PLAINTIFFS' REVERSE PASSING OFF CLAIM IS BARRED UNDER *DASTAR*. 23

VI. PLAINTIFFS' PRIVACY CLAIMS AGAINST GITHUB FAIL. 23

VII. PLAINTIFFS' UNJUST ENRICHMENT, CONSPIRACY, AND DECLARATORY RELIEF ALLEGATIONS FAIL ALONG WITH THE UNDERLYING CLAIMS. 25

CONCLUSION 25

TABLE OF AUTHORITIES

Page(s)

Cases

Apple Computer, Inc. v. Microsoft Corp.,
35 F.3d 1435 (9th Cir. 1994).. 11

Applied Equip. Corp. v. Litton Saudi Arabia Ltd.,
7 Cal. 4th 503 (1994) 25

Ashcroft v. Iqbal,
556 U.S. 662 (2009) 8, 21, 23

Author's Guild v. Google, Inc.,
804 F.3d 202 (2d Cir. 2015).. 6

Bed, Bath & Beyond of La Jolla, Inc. v. La Jolla Vill. Square Venture Partners,
52 Cal. App. 4th 867 (1997)... 20

Bosinger v. Belden CDT, Inc.,
358 F. App'x 812 (9th Cir. 2009) 25

Bounce Exchange, Inc. v. Zeus Enter. Ltd.,
No. 15cv3268, 2015 WL 8579023 (S.D.N.Y. Dec. 9, 2015).. 14

Cork v. CC-Palo Alto, Inc.,
534 F. Supp. 3d 1156 (N.D. Cal. 2021) 25

Crafty Prods., Inc. v. Fuqing Sanxing Crafts Co.,
839 F. App'x 95 (9th Cir. 2020) 19

Dastar Corp. v. Twentieth Century Fox Film Corp.,
539 U.S. 23 (2003) 23

Davis v. Federal Election Comm'n,
554 U.S. 724 (2008) 6

Design Basics, LLC v. WK Olson Architects, Inc.,
No. 17 C 7432, 2019 WL 527535 (N.D. Ill. Feb. 11, 2019) 14

Destfino v. Reiswig,
630 F.3d 952 (9th Cir. 2011)... 12

Doe v. Kamehameha Sch./Bernice Pauahi Bishop Est.,
596 F.3d 1036 (9th Cir. 2010).. 8

Does v. Advanced Textile Corp.,
214 F.3d 1058 (9th Cir. 2000).. 7, 8

Falkner v. Gen. Motors LLC,
393 F. Supp. 3d 927 (C.D. Cal. 2018).. 13, 14

Faulkner Press, L.L.C. v. Class Notes, L.L.C.,
756 F. Supp. 2d 1352 (N.D. Fl. 2010) 14

Frost-Tsuji Architects v. Highway Inn, Inc.,
No. CIV. 13-00496 SOM, 2015 WL 263556 (D. Haw. Jan. 21, 2015), *aff'd*,
700 F. App'x 674 (9th Cir. 2017) 14

Giddings v. Vison House Prod., Inc.,
No. CV 05-2963, 2007 WL 2274800 (D. Ariz. Aug. 7, 2007) 20

Google LLC v. Oracle Am., Inc.,
141 S. Ct. 1183 (2021).. 6, 11, 19

Harper & Row Publishers, Inc. v. Nation Enters.,
723 F.2d 195 (2d Cir. 1983).. 19

Harrington v. Pinterest, Inc.,
No. 20-cv-05290, 2021 WL 4033031 (N.D. Cal. Sept. 3, 2021) 10

Hobbs v. Sprague,
87 F. Supp. 2d 1007 (N.D. Cal. 2000) 25

I.C. v. Zynga, Inc.,
600 F. Supp. 3d 1034 (N.D. Cal. 2022) 8

Idema v. Dreamworks, Inc.,
162 F. Supp. 2d 1129 (C.D. Cal. 2001).. 19

In re iPhone Application Litig.,
No. 11-MD-02250, 2011 WL 4403963 (N.D. Cal. Sept. 20, 2011) 12

Ixchel Pharma, LLC v. Biogen, Inc.,
9 Cal. 5th 1130 (2020) 20, 21

Jacobsen v. Katzer,
535 F.3d 1373 (Fed. Cir. 2008).. 11

Kelly v. Arriba Soft Corp.,
336 F.3d 811 (9th Cir. 2003).. 6

Kelly v. Arriba Soft Corp.,
77 F. Supp. 2d 1116 (C.D. Cal. 1999).. 10, 14

Kodadek v. MTV Networks, Inc.,
152 F.3d 1209 (9th Cir. 1998).. 20

Korea Supply Co. v. Lockheed Martin Corp.,
29 Cal. 4th 1134 (2003) 21

Kwikset Corp. v. Super. Ct.,
51 Cal. 4th 310 (2011) 23

Lazar v. Superior Ct.,
12 Cal. 4th 631 (1996) 22

Logan v. Meta Platforms, Inc.,
No. 22-cv-01847, 2022 WL 14813836 (N.D. Cal. Oct. 25, 2022) 15

Lopez v. Bank of Am., N.A.,
505 F. Supp. 3d 961 (N.D. Cal. 2020) 23

Low v. LinkedIn Corp.,
900 F. Supp. 2d 1010 (N.D. Cal. 2012) 15, 16, 18

Lujan v. Defenders of Wildlife,
504 U.S. 555 (1992) 6, 7, 25

Maloney v. T3Media, Inc.,
853 F.3d 1004 (9th Cir. 2017) 18, 20

McGee v. S-L Snacks Nat'l,
982 F.3d 700 (9th Cir. 2020) 5, 6, 23

Melchior v. New Line Prods., Inc.,
106 Cal. App. 4th 779 (2003) 25

Mills v. Netflix, Inc.,
No. CV 19-7618, 2020 WL 548558 (C.D. Cal. Feb. 3, 2020) 13

Montz v. Pilgrim Films & Television, Inc.,
649 F.3d 975 (9th Cir. 2011) (en banc) 19

Murphy v. Millennium Radio Grp. LLC,
650 F.3d 295 (3d Cir. 2011) 14

Orchard Supply Hardware LLC v. Home Depot USA, Inc.,
967 F. Supp. 2d 1347 (N.D. Cal. 2013) 17

Perfect 10, Inc. v. Amazon.com, Inc.,
508 F.3d 1146 (9th Cir. 2007) 6

Petroliam Nasional Berhad v. GoDaddy.com, Inc.,
737 F.3d 546 (9th Cir. 2013) 13

Philpot v. Alternet Media, Inc.,
No. 18-cv-4479, 2018 WL 6267876 (N.D. Cal. Nov. 30, 2018) 10

Robinson Helicopter Co. v. Dana Corp.,
34 Cal. 4th 979 (2004) 22

Rosen v. Uber Techs., Inc.,
164 F. Supp. 3d 1165 (N.D. Cal. 2016) 21

Sega Enters. Ltd. v. Accolade, Inc.,
977 F.2d 1510 (9th Cir. 1992).......... 6

SellPoolSuppliesOnline.com, LLC v. Ugly Pools Ariz., Inc.,
804 F. App’x 668 (9th Cir. 2020) 15

Sony Corp. of Am. v. Universal City Studios, Inc.,
464 U.S. 417 (1984) 11

Stevens v. Corelogic, Inc.,
899 F.3d 666 (9th Cir. 2018).......... 9, 10, 11, 12

Sun Microsystems, Inc. v. Microsoft Corp.,
188 F.3d 1115 (9th Cir. 1999).......... 11

Swartz v. KPMG LLP,
476 F.3d 756 (9th Cir. 2007).......... 22

Sybersound Records, Inc. v. UAV Corp.,
517 F.3d 1137 (9th Cir. 2008).......... 23

Thunder Power New Energy Vehicle Dev. Co. v. Byton N. Am. Corp.,
340 F. Supp. 3d 922 (N.D. Cal. 2018) 8

TransUnion LLC v. Ramirez,
141 S. Ct. 2190 (2021) 5, 8

United States v. Bestfoods,
524 U.S. 51 (1998) 12

United States v. Doe,
655 F.2d 920 (9th Cir. 1980).......... 8

United States v. Elcom Ltd.,
203 F. Supp. 2d 1111 (N.D. Cal. 2002) 9

Warren v. Fox Family Worldwide, Inc.,
328 F.3d 1136 (9th Cir. 2003).......... 5

Westside Ctr. Assocs. v. Safeway Stores 23, Inc.,
42 Cal. App. 4th 507 (1996).......... 21

Young v. Facebook, Inc.,
790 F. Supp. 2d 1110 (N.D. Cal. 2011) 16

Statutes

Copyright Act, 17 U.S.C. §§ 101 *et seq.*

- § 102 18
- § 103 18
- § 106 18
- § 301(a) 18, 19
- § 411 6
- § 1202 .... 10, 11, 14
- § 1202(a) 10, 15, 16
- § 1202(b) 10, 11, 14, 15, 16
- § 1202(c) 10, 16

Cal. Bus. & Prof. Code §§ 17200 *et seq.* 23

California Civil Code

- § 1798.81.5(d) 25
- § 1798.150 .... 24, 25

Other Authorities

Federal Rules of Civil Procedure

- Rule 8 .... 12, 16
- Rule 9 .... 21, 22
- Rule 10 .... 3, 7
- Rule 12(b)(1) 3, 5
- Rule 12(b)(6) 7, 8

Webster's Third New International Dictionary (1993) 13

NOTICE OF MOTIONS AND MOTIONS

TO PLAINTIFFS AND THEIR ATTORNEYS OF RECORD:

PLEASE TAKE NOTICE THAT on May 4, 2023 at 2:00 p.m., before the Honorable Jon S. Tigar seated in Courtroom 6 of the United States Courthouse at Oakland, California, with appearances to be made by Zoom videoconference unless otherwise ordered by the Court, Defendants GitHub, Inc. ("GitHub") and Microsoft Corporation ("Microsoft") will, and hereby do, move, pursuant to Federal Rules of Civil Procedure 9, 10, 12(b)(1) and 12(b)(6), to dismiss the operative Complaint in these consolidated actions, filed as ECF No. 1 in the 22-cv-07074 action (the "Complaint" or "Compl."), in its entirety as to both GitHub and Microsoft. *See* ECF No. 47 (ordering consolidation and designation of the operative Complaint).

The grounds for the Motions are as follows. First, the Complaint should be dismissed in its entirety pursuant to Rules 10 and 12(b)(1) because Plaintiffs fail to identify themselves and fail to allege that they suffered any actual or threatened injury at the hands of either GitHub or Microsoft sufficient to give rise to a concrete controversy to be adjudicated by this Court. Second, the Complaint should be dismissed in its entirety pursuant to Rule 12(b)(6) (and Rule 9 with respect to Count IV for Fraud) because each and every one of the numerous claims for relief interposed against GitHub and Microsoft are either (a) legally foreclosed, (b) deficient because factual allegations required to support necessary elements are missing or implausible, or (c) both.

MEMORANDUM OF POINTS AND AUTHORITIES

INTRODUCTION AND SUMMARY OF ISSUES

Since the advent of open source software, developers have been free—indeed, encouraged—to use, study, change, and share source code for the broader good. It's an extraordinary idea, one that has fostered an immense body of public knowledge, ever-evolving and ever-available for the next generation of developers to build, collaborate, and progress. The open source model is essential to collaborative and communal software development. GitHub was founded on the basis of these ideals, and when Microsoft invested billions to acquire GitHub in 2018, it cemented its commitment to them. The transformative technology at issue in this case, Copilot, reflects GitHub and Microsoft's ongoing dedication and commitment to this profound human project. Copilot is a coding assistant tool that crystallizes the knowledge gained from billions of lines of public code, harnessing the collective power of open source software and putting it at every developer's fingertips.

Two anonymous detractors bring this lawsuit, claiming "software piracy on an unprecedented scale," and seeking penalties in excess of $9 billion. Despite that rhetoric, they do not advance a copyright infringement claim at all—doubtless an attempt to evade the limitations on the scope of software copyright and the progress-protective doctrine of fair use. Copilot withdraws nothing from the body of open source code available to the public. Rather, Copilot helps developers write code by generating suggestions based on what it has learned from the entire body of knowledge gleaned from public code. In so doing, Copilot advances the very values of learning, understanding, and collaboration that animate the open source ethic. With their demand for an injunction and a multi-billion dollar windfall in connection with software that they willingly share as open source, it is Plaintiffs who seek to undermine those open source principles and to stop significant advancements in collaboration and progress.

Plaintiffs' Complaint fails on two intrinsic defects: lack of injury and lack of an otherwise viable claim. Plaintiffs do not allege that this extraordinary new tool has harmed them in any way. They do not explain how teaching Copilot about their code has taken anything from them. They do not allege that Copilot has done anything improper with *their* code, nor that GitHub or

Microsoft did anything improper with *their* personal information. They seem to say that Copilot could theoretically suggest a snippet of code that matches something they have published, and do so without giving them proper attribution. How that series of hypothetical events could harm them is also unexplained. There is no case or controversy here—only an artificial lawsuit brought by anonymous Plaintiffs built on a remote possibility that they will fail to be associated with an AI-generated code snippet that could in theory be connected to copyrightable aspects of their source code. The Complaint's failure to make out a case for actual injury to named people requires dismissal under Rules 10 and 12(b)(1). Part I, *infra*.

Additionally, Plaintiffs do not state a claim. Their Complaint cycles through twelve purported counts, some of which embrace multiple theories, in search of a hook for their abstract grievance against Copilot. Despite their attempt to make a federal case out of it, none of these counts plausibly establishes that training or using Copilot violates any legal right. Part II, *infra*.

ALLEGATIONS OF THE OPERATIVE COMPLAINT

A. OpenAI Develops A Generative AI Tool Called Codex.

OpenAI is a nonprofit organization that develops machine learning models, also referred to as "Artificial Intelligence." Compl. ¶¶ 2, 126, 128. Such models are typically trained through exposure to a corpus of material called "training data." Compl. ¶ 79. The patterns discerned from the set of training data become part of the model, which can then generate answers based upon those patterns in response to user prompts. Compl. ¶ 81; *see* Compl. ¶ 52.

The model at issue in this case is called Codex. Codex is a generative AI model trained on publicly available computer source code. Compl. ¶¶ 52, 135. The model embodies "inferred … statistical patterns governing the structure of code," Compl. ¶ 52, which it has discerned from the training data based on "a complex probabilistic process." Compl. ¶ 79. It is thus capable, in response to a prompt, of "predic[ting] … the most likely [coding] solution." Compl. ¶ 79. "[E]ssentially [it] returns the solution it has found in the most [coding] projects when those projects are somehow weighted to adjust for whatever variables [the model] ha[s] identified as relevant." Compl. ¶ 79.

Generative AI models are capable of "simulat[ing] human reasoning or inference," engaging in the same sort of pattern recognition, synthesis, and prediction we do. 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. Compl. ¶ 81. At the same time, Codex "does not understand the meaning of code," Compl. ¶ 54, nor its "semantics and context the way humans do," Compl. ¶ 81. 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.

B. GitHub Offers Copilot, A Code Completion Tool Built On Codex.

GitHub Copilot is a programming assistant. Compl. at 8 n.4, ¶¶ 82-83. It "uses the OpenAI Codex to suggest code and entire functions in real-time" to software developers. Compl. ¶ 47. To use Copilot, a GitHub user installs it "as an extension to various code editors, including Microsoft's Visual Studio and VS Code." Compl. ¶ 67. "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. Compl. ¶ 67. Copilot is a subscription tool available to GitHub users for $10 per month or $100 per year. Compl. ¶ 8.

The version of Codex that powers Copilot was trained on billions of lines of code that GitHub users stored in public GitHub repositories. *See* Compl. ¶ 143. When GitHub users put their code on GitHub, they choose whether to make the code repositories private or public. Compl. ¶ 119. 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." Compl. Ex. 1 at 28 (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. Compl. Ex. 1 at 27-28 (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. Compl. ¶ 34 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* Compl. ¶ 34 n.4 & Appx. A.

C. Plaintiffs Sue Based On An Attribution Theory.

Plaintiffs are two anonymous GitHub users. Compl. ¶¶ 19-20. They both claim to have (at an unspecified time) "published Licensed Materials they owned a copyright interest in to at least one GitHub repository under one of the Suggested licenses." Compl. ¶¶ 19-20. But they do not allege, either expressly or on information and belief, that their Licensed Materials were used to train Codex or Copilot. The most charitable reading of the Complaint is that because their Licensed Materials were in public repositories, and since public repositories were used to train Codex and Copilot, Compl. ¶ 82, they believe their Licensed Materials were used to do so. The Complaint also is not clear on whether or to what extent the training of Copilot forms the basis of the Plaintiffs' claims for relief.

The crux of the Complaint, instead, is Copilot's suggestions, which Plaintiffs term "Output." Plaintiffs allege that these suggestions may sometimes match snippets of code used to train Copilot, but without providing information like authorship or licensing status. According to Plaintiffs, Copilot's "Output is often a near-identical reproduction of code from the training data," Compl. ¶ 46, and Copilot "has not been trained to provide Attribution." Compl. ¶ 56. On this basis, Plaintiffs allege that "Defendants stripped Plaintiffs' and the Class's attribution, copyright notice, and license terms from their code." Compl. ¶¶ 10-11, 78-81. Beyond that conclusory allegation, however, Plaintiffs do not connect their own code to any such Outputs. They identify no Output that has matched any of Plaintiffs' Licensed Materials. They identify no prompt that might produce such a match. They allege only a few examples of code Outputs matching *other* authors' code that have nothing to do with Plaintiffs. Compl. ¶¶ 48-81. And, while they assert that Copilot's suggestions will "often" match existing code, they point only to a study suggesting that "about 1% of the time, a suggestion … may contain some code snippets longer than ~150 characters that matches" some preexisting code. Compl. ¶ 90. Plaintiffs' allegations concerning

Copilot do nothing to connect their own code to an Output.

The Complaint nominally advances a dozen claims for relief related to Codex and Copilot. The first claim encompasses multiple types of alleged violations and both direct and secondary theories of liability. Compl. ¶¶ 142-71 (Count I). Other claims specify a remedy with multiple legal theories, or otherwise rely on multiple theories. Compl. ¶¶ 204-10 (Count VI); *id.* ¶¶ 211-14 (Count VII). Plaintiffs also tack on a theory of "civil conspiracy" covering most of the claims. Compl. ¶¶ 240-44.

But this long and complex Complaint speaks loudest with what it doesn't say. Plaintiffs fail to identify any of the "Licensed Materials" they allegedly placed in a GitHub public repository that reflect purported "copyright interests," or to tell us anything at all about those materials. The Complaint nowhere identifies any copyrighted work owned by either of the Plaintiffs, or any registration of such work. The Complaint fails to identify any use of their Licensed Materials. Although the case is supposedly about "software piracy on an unprecedented scale," Plaintiffs make no copyright infringement claim. And, Plaintiffs identify no personal identifying information that they stored in their public repositories on GitHub, or say how it was allegedly exposed by Codex or Copilot.

ARGUMENT

I. PLAINTIFFS LACK ARTICLE III STANDING AND THEREFORE SUBJECT MATTER JURISDICTION BECAUSE THEY HAVE NOT ALLEGED ACTUAL OR THREATENED INJURY.

GitHub and Microsoft move to dismiss under Rule 12(b)(1) because Plaintiffs fail to plead the actual or threatened injury necessary for Article III standing. If a plaintiff lacks Article III standing, the federal court lacks subject matter jurisdiction and the suit must be dismissed. *Warren v. Fox Family Worldwide, Inc.*, 328 F.3d 1136, 1140 (9th Cir. 2003). "'[A]t the pleading stage, the plaintiff must clearly allege facts demonstrating each element of standing.'" *McGee v. S-L Snacks Nat'l*, 982 F.3d 700, 705 (9th Cir. 2020) (cleaned up). One such element "plaintiffs must demonstrate [is] that they suffered a concrete harm." *TransUnion LLC v. Ramirez*, 141 S. Ct. 2190, 2200 (2021). This "injury in fact" requirement demands "more than an injury to a cognizable interest. It requires that the part[ies] seeking review be [themselves] among the injured." *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 563 (1992). A plaintiff must therefore allege an injury "that affects the plaintiff in a personal and individual way." *McGee*, 982 F.3d at 705 (cleaned up). And it must do so "for each claim … and for each form of relief that is sought." *Davis v. Federal Election Comm'n*, 554 U.S. 724, 734 (2008) (cleaned up).

The two anonymous Plaintiffs in this action, J. Does 3 and 4, do not meet this standard as to either of the Complaint's inconsistent theories of harm. The first theory is that Plaintiffs should have been identified in connection with their code (lack of attribution), and the second is that they should *not* have been identified (privacy). Yet the Complaint lacks allegations sufficient to demonstrate that Plaintiffs were or will be injured in either respect.

A. Plaintiffs' Lack-Of-Attribution Theory Is Insufficient To Confer Standing.

Both of the Plaintiffs allege that they are GitHub users who have published unspecified source code subject to unspecified open source licenses at an unspecified time. *See* Compl. ¶¶ 19-20. For all their references to "copyright interests" and a "Brave New World of Software Piracy," neither of the Plaintiffs allege ownership or infringement of any copyrighted (let alone registered, *see* 17 U.S.C. § 411) work. They do not even *identify* a copyrighted work. Plaintiffs thus allege no invasion of their copyright interests—an allegation that would run headlong into the doctrine of fair use. *See Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183 (2021); *Author's Guild v. Google, Inc.* 804 F.3d 202 (2d Cir. 2015); *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007); *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003); *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992).

Nor do Plaintiffs identify any harm that has come from the bare use of the contents of public repositories to train Codex or Copilot. Plaintiffs admittedly chose to make their source code freely available for inspection by anyone. Plaintiffs assert no legal interest that would entitle them to restrict the study by human or machine of freely available code. Plaintiffs do not explain how tangible harm to them results from an AI model's study of such code for training purposes. And the lack of such an allegation is no surprise, since the First Amendment and fair use generally protect such study. The open source principles embodied in common license agreements embrace learning and understanding from published code.

Plaintiffs instead conjure a theory of harm related not to copyright interests or training, but to lack of attribution. The apparent contention is that Copilot may generate suggested snippets of code that match snippets in existing GitHub projects, without providing credit. As explained below in connection with Microsoft and GitHub's Rule 12(b)(6) motion, Plaintiffs fail to state a viable claim based on this theory. But even assuming such a claim is conceivable in the abstract, there is no factual allegation in the Complaint suggesting that *these Plaintiffs* have suffered injury under their snippet-without-credit theory. They allege a few examples of a Copilot suggestion matching *someone else's* code. Compl. ¶¶ 66-76. They perform a spurious back-of-the-envelope calculation suggesting 12,000 users whose code might be matched by a Copilot suggestion. Compl. ¶ 91. Nothing in the Complaint, however, suggests that either of the Plaintiffs have been or will be among those users. That leaves no one in this case claiming that Copilot has actually caused them any harm at all—let alone pleading a *legal claim* based on that harm.

Plaintiffs also fail to allege facts supporting a reasonable inference that their code plausibly would become a Copilot Output. What code do they claim an interest in? What problem does that code solve? Where and how frequently has their published code been replicated by others such that Copilot would identify it as "the most likely solution to a given prompt," Compl. ¶ 79? What prompt could potentially generate a match, and why is it likely that a user would enter this prompt? Without these types of allegations, Plaintiffs allege at best the sort of "conjectural or hypothetical" injury that cannot confer standing. *Lujan*, 504 U.S. at 650.

Indeed, Plaintiffs' failure to identify *themselves* in violation of Rule 10 underscores the wholly speculative nature of the case. The "use of fictitious names runs afoul of the public's common law right of access to judicial proceedings" and is counter to "Rule 10(a)'s command that the title of every complaint 'include the names of all the parties.'" *Does v. Advanced Textile Corp.*, 214 F.3d 1058, 1067 (9th Cir. 2000) (quoting Fed. R. Civ. P. 10(a)).[*] If what Plaintiffs truly sought was attribution for their code, one would expect them to identify themselves and that code. Instead, they offer only abstraction.

---

[*] Plaintiffs have not obtained leave to proceed without identifying themselves, nor do they allege facts that would meet one or more of three scenarios justifying a use of pseudonyms: (1) where the anonymous party faces retaliation, (2) where the litigation concerns a "sensitive and highly personal" matter, and (3) where the anonymous party admits an intention to engage in illegal conduct. *Advanced Textile*, 214 F.3d. at 1068; *see generally Doe v. Kamehameha Sch./Bernice Pauahi Bishop Est.*, 596 F.3d 1036, 1043-44 (9th Cir. 2010) (noting risk of "physical harm" or

B. Plaintiffs Also Do Not Allege Privacy-Based Injury.

Even as Plaintiffs hypothesize harm from failing to be identified, they also allege their privacy is injured through misuse of their personal identifying information ("PII"). Compl. ¶¶ 225-39. This second and inconsistent theory of injury, like the first, fails to demonstrate any actual or threatened harm because Plaintiffs have alleged no details concerning collection or mishandling of any of their PII. They have not identified the type of information at issue, any specific information of their own, how and from where any Defendant allegedly collected that information, or how that information was mishandled.

Plaintiffs therefore cannot allege a privacy-based injury, like invasion of privacy, loss of control over private information, or disclosure of private information. *Cf. TransUnion*, 141 S. Ct. at 2204 (identifying types of harms supporting standing). Indeed, any such injury appears *implausible* based on what is alleged: if Plaintiffs are claiming that their PII was collected from *public* repositories on GitHub, it is difficult to see how collection of that information could possibly effect a *privacy* harm. In any event, Plaintiffs' failure to allege privacy-based injury requires dismissal. *E.g.*, *I.C. v. Zynga, Inc.*, 600 F. Supp. 3d 1034, 1048-50 (N.D. Cal. 2022).

II. PLAINTIFFS FAIL TO STATE A DMCA CLAIM.

Apart from failing to establish standing, Plaintiffs also do not state a claim. To survive a motion to dismiss under Rule 12(b)(6), a complaint must contain factual matter that, when accepted as true, states a claim that is plausible on its face. *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009). If a claim lacks a cognizable legal theory or lacks sufficient facts to plausibly support a cognizable legal theory, dismissal is warranted. *Thunder Power New Energy Vehicle Dev. Co. v. Byton N. Am. Corp.*, 340 F. Supp. 3d 922, 925 (N.D. Cal. 2018).

Count I alleges violations of 17 U.S.C. § 1202, a part of the Digital Millennium Copyright Act ("DMCA"). The DMCA was enacted to balance the rights of users with those of copyright

---

"severe threats of deportation, arrest, and imprisonment" as justifying anonymity); *United States v. Doe*, 655 F.2d 920, 922 n.1 (9th Cir. 1980) (pseudonyms permitted in cases concerning abortion and contraceptive bans).

owners in light of new technologies. *See generally United States v. Elcom Ltd.*, 203 F. Supp. 2d 1111 (N.D. Cal. 2002). Section 1202 is directed to the potential for bad acts that facilitate mass infringement in the electronic age. It protects the "[i]ntegrity of copyright management information" ("CMI")—that is, certain categories of "information conveyed in connection with copies … of a work," like titles or authorship information. 17 U.S.C. § 1202(c). Section 1202 proscribes the provision or distribution of false CMI, intentional removal of CMI, alteration of CMI, and distribution of works with removed or altered CMI. Consistent with its focus on thwarting piracy, each subsection contains a double-scienter requirement. *See id.* A defendant is liable only if it had both the requisite scienter with respect to the violative conduct (i.e., provision of false CMI, removal of CMI, etc.) *and* some degree of knowledge that this conduct would "induce, enable, facilitate, or conceal infringement." *Id.* § 1202(a)-(b). Neutral technology like Copilot cannot satisfy that standard. Part A, *infra*. Even were that possible as a theoretical legal matter, Count I lumps far too much together to provide each of GitHub and Microsoft fair notice of who is alleged to have done what to violate § 1202. Part B, *infra*. Plaintiffs' purported claims under each subsection fail for multiple additional reasons. Sections C-D, *infra*.

A. Plaintiffs Cannot Allege The Required Likelihood Of Infringement.

All of the § 1202 claims fail because Plaintiffs do not allege any likelihood that Copilot will result in copyright infringement of any of their works, as required to state a claim under *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 673 (9th Cir. 2018). To plead a claim under § 1202(a) pertaining to false CMI, a plaintiff must allege that the defendant acted "with intent to induce, enable, facilitate, or conceal infringement." A claim under § 1202(b) regarding altered or removed CMI similarly requires that a defendant either "have knowledge" or have "reasonable grounds to know" of the same result. Of necessity, then, the tampering with CMI contemplated by § 1202 must be part of a scheme involving purposeful infringement.

That line is critical in a case like this one, where a plaintiff attacks a dual-use technology—a system that is not aimed at infringement, but may nevertheless touch or concern CMI in some way. Whether or not they advance copyright infringement claims, Plaintiffs must still explain in the context of their § 1202 claims how infringement is occurring or is likely to occur. They cannot. At most, Plaintiffs allege Copilot's suggestions will match snippets of existing code merely 1% of the time. Compl. ¶ 90. And even within that miniscule percentage, they do not try to allege *copyright infringement*—a virtual impossibility anyway given the various copyright-based obstacles (originality, fair use, etc.) that would preclude a claim. Plaintiffs do not allege a likelihood of infringement, let alone *purposeful* infringement.

Stevens addressed this scenario in a case involving § 1202(b)'s "reasonable grounds to know" requirement—the lowest of § 1202's scienter requirements. Photographers challenged the defendant's software program, which "downsampled" (i.e., condensed the size of) images for use in electronic real estate listings, and in the process failed to preserve the CMI metadata from the plaintiffs' photographs. 899 F.3d at 671. The plaintiffs nonetheless argued that they satisfied the "reasonable grounds to know" requirement because when CMI is removed, "one method of identifying an infringing photograph has been impaired, [such that] someone *might* be able to use their photographs undetected." *Id.* at 673.

The Ninth Circuit rejected the argument. Although a § 1202 plaintiff "need not show that any specific infringement has already occurred," it still "must make an affirmative showing … that the defendant was aware or had reasonable grounds to be aware of the probable future impact of its actions." *Id.* at 674. And this requires a showing "from which one can infer that future infringement is *likely* … to occur as a result of the removal or alteration of CMI." *Id.* at 675 (emphasis added). The rule makes good sense: If infringement is merely a "general possibility," *id.* at 673, there is no plausible inference that the defendant's conduct was done with the requisite mental state. *See also Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999) (rejecting DMCA claim based on "unintended side effect" of web crawler's operation), *rev'd on other grounds*, 336 F.3d 811 (9th Cir. 2003). Courts have dismissed § 1202 claims at the pleading stage for failing to satisfy this requirement. *See Harrington v. Pinterest, Inc.*, No. 20-cv-05290, 2021 WL 4033031, at *6 (N.D. Cal. Sept. 3, 2021); *Philpot v. Alternet Media, Inc.*, No. 18-cv-4479, 2018 WL 6267876, at *5 (N.D. Cal. Nov. 30, 2018).

Much as in *Stevens*, Plaintiffs here challenge technology whose purpose is unrelated to infringement or CMI, but that incidentally may not preserve CMI. Plaintiffs allege that Copilot "has not been trained to provide Attribution," Compl. ¶ 56, and that, "[o]n information and belief, Defendants could have trained Copilot to include attribution, copyright notices, and license terms." Compl. ¶ 161. Plaintiffs thus allege that Copilot is CMI-agnostic. Affirmative steps would have to be taken to modify the technology to account for CMI. That is the precise scenario in which *Stevens* demands a showing that the technology's incidental effects with respect to CMI have a likelihood of causing, facilitating, or concealing infringement.

To overcome *Stevens*, Plaintiffs would need to allege facts showing that the technology at issue is likely to result in widespread infringement. Not only do Plaintiffs fail to allege such facts, but it is obvious from the Complaint that Copilot is not such a machine. Plaintiffs allege that suggestions match training data very rarely. Even assuming that were infringement, there is no obligation in copyright law to modify technology to avoid all possibility of infringement. *See Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417 (1984). Finally, and most importantly, it makes no sense to assume that any matching Outputs that might occur constitute copyright infringement. Even when a snippet matches training data, there are a host of additional hurdles to the notion that any particular Output would result in copyright infringement: the copyrightability of such material in light of doctrines of merger and scènes à faire, *see Apple Computer, Inc. v. Microsoft Corp.*, 35 F.3d 1435, 1444-45 (9th Cir. 1994); the requirement of "virtually identical" copying of thinly protected elements, *see id.* at 1444; the scope of authorization through licensing, *compare Sun Microsystems, Inc. v. Microsoft Corp.*, 188 F.3d 1115 (9th Cir. 1999), *with Jacobsen v. Katzer*, 535 F.3d 1373 (Fed. Cir. 2008); and fair use, *see Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183 (2021). Given these hurdles, Copilot is simply not the type of technology that could ever satisfy the requirements of *Stevens*.

Similarly, just as in *Stevens*, Plaintiffs fail to offer any "specific [allegations] that removal of CMI … will impair their policing of infringement." 899 F.3d at 675. They do not, for example, "aver" that they "have ever used CMI … to prevent or detect copyright infringement." *Id.* And just as "a party intent on using a copyrighted photograph undetected can itself remove any CMI metadata," *id.* at 676, a party intent on using copyrighted code without reproducing licensing information could readily do so. So, "one cannot plausibly say that removal [by Copilot] 'will'" make it easier to infringe Plaintiffs' works. *Id.*

It is Plaintiffs' prerogative to not assert a claim for copyright infringement. But *Stevens* requires plausible allegations that Copilot's operation with respect to CMI is likely to have that result. Plaintiffs make no such allegations, and the nature of Copilot precludes them from doing so. Count I should be dismissed with prejudice.

B. Count I Impermissibly Lumps Together Multiple Defendants, Claims, And Theories Of Liability.

Plaintiffs cannot get around *Stevens*, but in any event the "failure to allege what role each Defendant played in the alleged harm makes it exceedingly difficult, if not impossible, for individual Defendants to respond to [the] allegations." *In re iPhone Application Litig.*, No. 11-MD-02250, 2011 WL 4403963, at *8 (N.D. Cal. Sept. 20, 2011). Such a pleading therefore does not satisfy Rule 8(a). *Id.*; *see Destfino v. Reiswig*, 630 F.3d 952, 958-59 (9th Cir. 2011) (affirming dismissal of "shotgun pleading").

Count I is a textbook example of this problem. It is alleged against "All Defendants." Over the course of a nearly 30-paragraph thicket of assertions, Plaintiffs appear to claim violation of all five of the statute's prohibitions. Compl. ¶¶ 142-71. They also purport, by parenthetical under the title of "COUNT I," to advance "(Direct, Vicarious, and Contributory)" theories of liability. *Id.* at 34. No allegation is specific to either GitHub or Microsoft. Nothing elsewhere in the Complaint explains what any particular Defendant is supposed to have done.

This defect is worst with respect to Microsoft. Plaintiffs allege that Microsoft owns GitHub and is an investor in OpenAI. Compl. ¶¶ 5-7. But it is basic corporate law that absent some disregard of corporate formalities, Microsoft cannot be held liable as a corporate parent or shareholder. *See United States v. Bestfoods*, 524 U.S. 51, 61 (1998). As for Microsoft's subsidiary GitHub, there is no explanation of what GitHub supposedly did with respect to the alteration or removal of CMI. And since there is no allegation suggesting that GitHub or Microsoft engaged in primary conduct involving CMI, Plaintiffs presumably mean to sweep them in with the "vicarious" and "contributory" labels. But § 1202 does not provide for such theories of liability on its face, and it is doubtful that the Ninth Circuit would find § 1202's double-scienter requirement compatible with theories of secondary liability. *Cf. Petroliam Nasional Berhad v. GoDaddy.com, Inc.*, 737 F.3d 546 (9th Cir. 2013) (refusing extra-statutory contributory liability theory under anti-cybersquatting statute in part because of "bad faith" requirement).

Plaintiffs are claiming entitlement to statutory damages of over $9 billion for their § 1202 claims. Compl. at 52 n.42. Defendants cannot properly address these claims without knowing what Plaintiffs say each Defendant did to violate the statute.

C. Plaintiffs Have Not Plausibly Alleged That GitHub Or Microsoft Intentionally Or Knowingly Removes Or Alters CMI Under § 1202(b).

Plaintiffs' claims under § 1202(b)(1) and (3) fail for an independent reason. Section 1202(b)(1) prohibits a defendant from "intentionally remov[ing] or alter[ing] [CMI]," *see* Compl. ¶ 148, while § 1202(b)(3) bars "distribut[ing] … copies of works … knowing that [CMI] has been removed or altered," *see* Compl. ¶ 154. These provisions require the active step of removal or alteration of CMI in connection with an identical copy of the work. But all the Complaint alleges is the passive non-inclusion of CMI with a proffered snippet that may incidentally match a portion of a work. That is not enough.

This result flows directly from § 1202's text. The terms "remove" and "alter" both connote active conduct. To remove is "to move by lifting, pushing aside, or taking away or off" or "to get rid of." Webster's Third New International Dictionary 1921 (1993). To alter is "to cause to become different." *Id.* at 63. As explained above, CMI is defined as information "conveyed in connection *with copies* … of a work"—that is, in or directly connected with "the body of or the actual work itself." *See Mills v. Netflix, Inc.*, No. CV 19-7618, 2020 WL 548558, at *3 (C.D. Cal. Feb. 3, 2020). To remove or alter CMI, then, is to engage in active conduct to get rid of or change CMI that would otherwise naturally accompany a copy of the work from which the CMI is allegedly removed or altered. *See Falkner v. Gen. Motors LLC*, 393 F. Supp. 3d 927, 938-39 (C.D. Cal. 2018).

Courts have consistently held this line. They have rejected § 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 SOM, 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*, 77 F. Supp. 2d at 1122 (thumbnail versions of images). Courts have recognized § 1202(b) claims only for active conduct in connection with identical copies of a work—like a defendant cropping a "gutter credit" from an otherwise identical photo, *e.g.*, *Murphy v. Millennium Radio Grp. LLC*, 650 F.3d 295 (3d Cir. 2011), or deleting CMI *from within* identical content that was excerpted. *E.g.*, *Bounce Exchange, Inc. v. Zeus Enter. Ltd.*, No. 15cv3268, 2015 WL 8579023, at *2-3 (S.D.N.Y. Dec. 9, 2015) (CMI actively removed from lines of software code).

Plaintiffs allege only that Copilot "has not been trained" to actively identify and reproduce any CMI that may or may not be present. Compl. ¶ 56. And rather than alleging removal of CMI connected with an identical copy of any work, their claims are based on mere code suggestions that *may* match small snippets of code contained in other works. Indeed, as Plaintiffs acknowledge, Copilot's suggestions are not copied from any particular source, but generated based on patterns it has "seen *the most*," Compl. ¶ 73. Plaintiffs therefore cannot state a claim that CMI is actively removed from any work.

D. Plaintiffs Have Not Plausibly Alleged That GitHub Or Microsoft Provides, Distributes, Or Imports For Distribution Any False CMI.

In addition to the failure to plead likely copyright infringement, Section B, *supra*, Plaintiffs' claims under § 1202(a)(1)-(2) and (b)(2) also fail for an independent reason. Those provisions bar a defendant from "provid[ing] [CMI] that is false," § 1202(a)(1), *see* Compl. ¶ 158; "distribut[ing] or import[ing] for distribution [CMI] that is false," § 1202(a)(2), Compl. ¶ 158; and "distribut[ing] or import[ing] for distribution [CMI] knowing that the [CMI] has been removed or altered," § 1202(b)(2), Compl. ¶ 153. But Plaintiffs fail to plausibly allege that

GitHub or Microsoft provide, distribute, or import for distribution any CMI at all.

Plaintiffs' only allegation in this regard is that "Defendants have a business practice of asserting and/or implying that Copilot is the author of the Licensed Materials." Compl. ¶ 158. Plaintiffs provide no instance of such an assertion or implication, either as to works users place in public repositories or Copilot's suggestions. And GitHub's Terms of Service ("TOS"), attached to the Complaint, directly refute Plaintiffs' allegation: "GitHub does not claim any rights in [Copilot] Suggestions, and you retain ownership of and responsibility for Your Code, including Suggestions you include in Your Code." Compl. Ex. 1 at 38 (Copilot Additional Terms at 1).

In any event, Plaintiffs do not allege any distribution of CMI by GitHub or Microsoft *at all*, let alone false CMI. Section 1202(c) defines CMI as "information conveyed *in connection with copies* … of a work." (emphasis added). Courts have construed the "in connection with copies" language to require a specific, close nexus between the information conveyed and the copy of the work. For example, a generic copyright notice on a webpage that was not "located on or next to [the plaintiff's] photos" lacked sufficient connection. *SellPoolSuppliesOnline.com, LLC v. Ugly Pools Ariz., Inc.*, 804 F. App'x 668, 670-71 (9th Cir. 2020); *see Logan v. Meta Platforms, Inc.*, No. 22-cv-01847, 2022 WL 14813836, at *8 (N.D. Cal. Oct. 25, 2022). The Complaint identifies no information provided anywhere, let alone in connection with a copy of any work. As for their suggestion that GitHub or Microsoft "imply[]" authorship, implications are not "information" that can be provided, distributed, or imported for distribution. They are therefore not CMI at all. And even if they were, Plaintiffs point to no implication of authorship made in connection with a copy of a work. Plaintiffs' claims under § 1202(a) and (b)(2) should therefore be dismissed.

III. PLAINTIFFS FAIL TO STATE A BREACH-OF-LICENSE CLAIM.

"Under California law, to state a claim for breach of contract a plaintiff must plead the contract, plaintiffs' performance (or excuse for nonperformance), defendant's breach, and damage to plaintiff therefrom." *Low v. LinkedIn Corp.*, 900 F. Supp. 2d 1010, 1028 (N.D. Cal. 2012) (quotation marks omitted). Count II fails this test.

A. Plaintiffs Fail To Specify The Contract Provisions Allegedly Breached.

Plaintiffs' breach of contract claim fails at the outset because the Complaint does not identify the contract(s) or provision(s) that each Defendant allegedly breached. *See Low*, 990 F. Supp. 2d at 1028. One would ordinarily see a contract claim alleging "plaintiff and defendant entered into X contract and then defendant violated paragraph Y of that contract by doing Z." None of that exists here. Instead, all Plaintiffs assert is that "Defendants [lumped together] violated the [unspecified] Licenses governing use of the [unidentified] Licensed Materials by using them to train Copilot." Compl. at 10. There is no identification of the who and how of the contractual relationships, no identification of the object(s) of any alleged contract, nor identification of how any particular act by any particular Defendant violated any particular contractual provision. Plaintiffs do not allege *how* training Copilot on public GitHub repositories breaches any provision of any contract. This is not just procedural theater. It is important *because none of the open source licenses attached to the Complaint appear to prohibit such activity in any way*. On the contrary, the principles embodied in customary open source licenses contemplate broad public rights to inspect, learn from, and build upon code.

Rule 8 requires at least a minimal effort in a contract claim to explain the breach. Count II is defective in this respect, and the failure to offer any explanation is particularly egregious with respect to Defendant Microsoft, which is not alleged to have done *anything* with Plaintiffs' "Licensed Materials." Because Plaintiffs fail to plead the "specific provisions in the contract … the defendant is said to have breached," *Young v. Facebook, Inc.*, 790 F. Supp. 2d 1110, 1117 (N.D. Cal. 2011), their claims should be dismissed.

B. To The Extent Count II Is Based On Training Copilot With Code Found In Public Repositories, It Is Foreclosed By GitHub's Terms Of Service.

In all events, to the extent Plaintiffs' unexplained breach theory against GitHub is predicated on training Copilot with code published in public GitHub repositories, Count II is foreclosed by GitHub's Terms of Service. Compl. Ex. 1.

Consistent with open source licensing principles, GitHub's TOS expressly authorizes the training of Copilot. The TOS informs users that "You own content you create, but you allow us certain rights to it." Compl. Ex. 1 at 27 (TOS at 6). "These license grants apply to Your Content" when uploaded to GitHub, notwithstanding any other license terms that might be attached. *Id.* Every GitHub user chooses whether to allow their "repositories to be viewed publicly." *Id.* No user is required to make a repository public. If they choose to do so, they "grant [GitHub] … the right to store, archive, parse, and display" publicly posted content and "make incidental copies, as necessary to provide the Service, including improving the Service over time." *Id.* This includes the right to "copy" public material "to our database," "parse it into a search index or otherwise analyze it on our servers," and "share it with other users." *Id.* at 27-28 (TOS at 6-7). The "Service" includes all "the applications, software, products, and services provided by GitHub," *id.* at 24 (TOS at 3), which includes Copilot. *See* Compl. ¶ 8; Compl. Ex. 1 at 24, 38 (TOS at 3, Copilot Additional Terms).

As Plaintiffs allege, Copilot functions by evaluating the code contained in GitHub's public repositories, Compl. ¶¶ 82-83, in order to "infer[] statistical patterns governing the structure of code." Compl. ¶ 52. 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." Compl. Ex. 1 at 27-28, 24 (TOS at 6-7, 3). Even if the 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 a complaint alleges facts that demonstrate 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.

C. Plaintiffs Have Not Plausibly Alleged Breach Based On An Output Theory.

Nor do Plaintiffs plausibly allege breach based on the use of Copilot. Plaintiffs allege that "when providing Output, Copilot does not comply with [open source license] terms." Compl. ¶ 174. But Plaintiffs have not alleged that Copilot has provided any of *their* code subject to an attribution requirement, nor plausibly pled that such a suggestion is likely. Nor do they explain how, absent that, there is a breach by any of these Defendants. So even if Defendants knew which provisions of which contracts were at issue, Plaintiffs have not plausibly alleged a breach or plausibly alleged any damages to them—both essential elements of their claim. *Low*, 900 F. Supp. 2d at 1028; *see also* Part I, *supra*.

Plaintiffs allege several examples of Copilot suggestions matching others' code. *See* Compl. ¶¶ 66-76; Compl. ¶ 87 ("code from the game Quake III"); Compl. ¶ 88 (unspecified "code that had been released under a license that allowed its use only for free games"); Compl. ¶ 89 ("code belonging to" a "Texas A&M computer-science professor"). None of these alleged licensors appear to be named Plaintiffs (and Plaintiffs would be hard pressed to claim that Microsoft—which owns Quake—is suing itself and its subsidiary in this lawsuit). Plaintiffs cannot save their claims with surmise that "if Licensed Materials have been posted to a GitHub public repository," they "can be reasonably certain [those materials are] sometimes returned to users as Output." Compl. ¶ 82. "1% of the time," Compl. ¶ 90, is hardly likely—and in any event, Plaintiffs fail to allege how even this 1% would breach any license.

IV. PLAINTIFFS' TORT AND UCL CLAIMS FAIL.

Plaintiffs advance claims against GitHub and Microsoft under California law for tortious interference with contract and economic advantage (Count III), against GitHub for fraud (Count IV), and against GitHub for unlawful competition under the UCL (Counts VI and VII). Each is preempted by the Copyright Act and deficient in other respects.

A. Plaintiffs' Tort And UCL Claims Are Preempted By The Copyright Act.

Section 301 of the Copyright Act preempts "legal or equitable rights that are equivalent to any of the exclusive rights within the general scope of copyright … and come within the subject matter of copyright." 17 U.S.C. § 301(a). Courts evaluate copyright preemption under a "two-part test," holding state law claims preempted where (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) "the rights asserted under state law are equivalent to the rights contained in 17 U.S.C. § 106." *Maloney v. T3Media, Inc.*, 853 F.3d 1004, 1010 (9th Cir. 2017) (quotation marks omitted).

The subject matter requirement of the test is plainly satisfied. Plaintiffs' tortious interference, fraud, and UCL claims all involve Plaintiffs' claimed copyright interests in computer code, which constitutes a "literary work" under § 102. *See Google*, 141 S. Ct. at 1196. Plaintiffs' tortious interference claim alleges that Defendants "improperly use[] Copilot to create Derivative Works," Compl. ¶ 189, the subject matter of § 103. The fraud claim likewise is based on the allegation that Defendants "sell[] … Licensed Materials as part of Copilot." Compl. ¶ 197. And aspects of Plaintiffs' UCL claims based on non-privacy state law causes of action likewise involve "code." Compl. ¶ 212(b). These claims therefore involve the "subject matter" of copyright, a result that obtains even if the code is not actually protected by the Copyright Act. *See Montz v. Pilgrim Films & Television, Inc.*, 649 F.3d 975, 979 (9th Cir. 2011) (en banc) ("[T]he scope of the subject matter of copyright law is broader than the protections it affords.").

The second step is also met with respect to each claim—indeed, the claims quote nearly verbatim from the Copyright Act's list of exclusive rights. As noted above, Plaintiffs' tortious interference claim is based on the allegation that Defendants "create Derivative works," Compl. ¶ 189, mirroring § 106(2)'s grant of the exclusive right "to prepare derivative works." And Plaintiffs allege no other conduct or element of their tortious interference claim that would render it "qualitatively different" from a claim of infringement of the right to create derivative works. *Crafty Prods., Inc. v. Fuqing Sanxing Crafts Co.*, 839 F. App'x 95, 98-99 (9th Cir. 2020) (tortious interference claim preempted); *Harper & Row Publishers, Inc. v. Nation Enters.*, 723 F.2d 195, 201 (2d Cir. 1983) (tortious interference claim preempted where "unauthorized publication is the gravamen of the[] claim"), *rev'd on other grounds*, 471 U.S. 539 (1985).

Similarly, Plaintiffs' fraud claim alleges merely that GitHub stated that it would not sell or distribute Plaintiffs' code, then "failed to honor [these] representations" by selling or distributing Plaintiffs' code. Compl. ¶¶ 195-97. This claim is indistinguishable from one for infringement of § 106(3)'s distribution right. Although fraud claims may escape preemption where they involved intentional misrepresentations tangentially related to the subject matter of copyright, Plaintiffs allege no misrepresentation at all—their claim is just an attempt to impose liability for distributing code. This disguised copyright claim is preempted. *See Idema v. Dreamworks, Inc.*, 162 F. Supp. 2d 1129, 1191-92 (C.D. Cal. 2001); *Giddings v. Vison House Prod., Inc.*, No. CV

05-2963, 2007 WL 2274800, at *3 (D. Ariz. Aug. 7, 2007).

To the extent Plaintiffs' UCL claim is predicated on the above 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). And the same analysis requires dismissal of Plaintiffs' invocation, under § 17200 and the common law, of the remedy of unjust enrichment, which once again is based on the allegation that Defendants "used [Plaintiffs' code] to create Derivative Works" and "res[old] [them]." Compl. ¶¶ 204-10.

B. Plaintiffs Do Not Plausibly Allege A Contract Or Business Expectancy, Nor Defendants' Knowledge Of One As Required For An Interference With Economic Advantage Claim.

Plaintiffs' tortious interference claim (Count III) against GitHub and Microsoft is deficient in multiple other ways. "Tortious interference with contractual relations requires (1) the existence of a valid contract between the plaintiff and a third party; (2) the defendant's knowledge of that contract; (3) the defendant's intentional acts designed to induce a breach or disruption of the contractual relationship; (4) actual breach or disruption of the contractual relationship; and (5) resulting damage." *Ixchel Pharma, LLC v. Biogen, Inc.*, 9 Cal. 5th 1130, 1141 (2020) (quotation marks omitted). The Complaint falls short on every element.

Although Plaintiffs label their claim one for tortious interference with *contract*, they identify no contractual relationship between themselves and anyone else—much less a contract that GitHub and Microsoft knew about. *See* Compl. ¶¶ 188-93; *see also Bed, Bath & Beyond of La Jolla, Inc. v. La Jolla Vill. Square Venture Partners*, 52 Cal. App. 4th 867, 878 (1997) (plaintiff must allege an existing, enforceable contract). While Plaintiffs allege that they deposited materials in public GitHub repositories subject to open source licenses, they do not allege that any third party ever entered into a contract with them in regard to those materials. Nor do Plaintiffs plausibly allege that GitHub or Microsoft committed "intentional acts designed to induce a breach or disruption" or "actual breach or disruption" with respect to an open source license granted by Plaintiffs to any unidentified third party. *Ixchel*, 9 Cal. 5th at 1141. All they allege is that "[a]t GitHub's upcoming yearly conference, GitHub Universe 2022, it will host a presentation called 'How to compete with open source—and win.'" Compl. ¶ 190. Plaintiffs construct their claim on a tortured interpretation of this title, as though the presentation provides a blueprint on how to compete *against* open source, when the more natural reading is that it is about how to compete effectively *using* open source. But in all events, this is hardly an allegation of a specific act to knowingly interfere with specific contracts. And in any event, "compet[ing]" and "win[ning]" is not a tort. *See Ixchel*, 9 Cal. 5th at 1142. The Complaint therefore states no claim for intentional interference with contract against GitHub or Microsoft.

The Complaint also fails to state a claim for the related tort of interference with prospective economic advantage. 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 knowingly interfered with an economic relationship between the plaintiff and some third party, which carries the probability of future economic benefit to the plaintiff." *Ixchel*, 9 Cal. 5th at 1141 (cleaned up). Specific known relationships must be identified. *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).

Here, Plaintiffs fail to allege any specific relationship with a reasonably likely economic benefit that GitHub and Microsoft knew about and intended to interfere with. The Complaint merely asserts that "Defendants have wrongfully interfered with the business interests and expectations of Plaintiffs and the Class by improperly using Copilot to create Derivative Works that compete against OSC." Compl. ¶ 189. This conclusory "the-defendant-unlawfully-harmed-me accusation," *Iqbal*, 556 U.S. at 678, does not identify the "economic relationship" supposedly interfered with or plausibly allege the "probability of future economic benefit" claimed to be disrupted. Nor does the Complaint allege "actual disruption" or specific "harm," two additional elements of the claim. *Korea Supply Co. v. Lockheed Martin Corp.*, 29 Cal. 4th 1134, 1153 (2003). As a result, the Complaint does not state a claim for any economic interference tort.

C. Plaintiffs' Fraud Claim Against GitHub Is Barred By The Economic Loss Rule And Fails Rule 9.

"The elements of fraud … are (a) misrepresentation (false representation, concealment, or nondisclosure); (b) knowledge of falsity (or 'scienter'); (c) intent to defraud, i.e., to induce reliance; (d) justifiable reliance; and (e) resulting damage." *Lazar v. Superior Ct.*, 12 Cal. 4th 631, 638 (1996) (quoting 5 Witkin, Summary of Cal. Law (9th ed. 1988) Torts, § 6765, p. 778). Federal Rule of Civil Procedure 9(b) further requires Plaintiffs to "state with particularity the circumstances constituting fraud." This required "specificity includ[es] an account of the time, place, and specific content of the false representations as well as the identities of the parties to the misrepresentations." *Swartz v. KPMG LLP*, 476 F.3d 756, 764 (9th Cir. 2007) (cleaned up).

Plaintiffs assert that "in both its Terms of Service and its Privacy Statement, GitHub promises not to sell Licensed Materials or anything else uploaded to or shared with GitHub." Compl. ¶ 195. But Plaintiffs nowhere allege that GitHub made these statements with "knowledge of falsity," "intent to defraud," or to "induce [unjustified] reliance." *Lazar*, 12 Cal. 4th at 638. The Complaint states merely that Plaintiffs "relied upon [GitHub's] representations in choosing to upload Licensed Materials to GitHub," in part because "GitHub has long held itself out as the best place to host open source code repositories." Compl. ¶ 196. These vague assertions do not suggest GitHub's intentions to commit fraud, particularly in light of the other provisions of the TOS. *See supra* 3, 15-17. Nor, as discussed, have Plaintiffs alleged any "resulting damage."

Ultimately, then, the basis of Plaintiffs' claim is merely that GitHub "failed to honor its representations." Compl. ¶ 197. That is a breach of contract claim, and thus barred by the economic loss rule. "The economic loss rule" dictates that "purely economic loss due to disappointed expectations" cannot be remedied in tort "unless [the Plaintiff] can demonstrate harm above and beyond a broken contractual promise," therefore "prevent[ing] the law of contract and the law of tort from dissolving one into the other." *Robinson Helicopter Co. v. Dana Corp.*, 34 Cal. 4th 979, 988 (2004) (cleaned up). The fraud claim against GitHub is merely derivative of alleged promises in the contractual TOS and the Complaint identifies no loss, let alone one independent of the promises of the TOS. Plaintiffs' fraud claims should be dismissed.

D. Plaintiffs Lack Standing To Bring A UCL Unlawful Competition Claim And Fail To State A Claim.

Plaintiffs duplicate their DMCA, tort, passing off, and privacy claims as purported violations by GitHub of California's Unfair Competition Law, Cal. Bus. & Prof. Code §§ 17200, *et seq.* (Count VII). This is dead on arrival. The "UCL's standing requirements" are even "more stringent than the federal standing requirements." *Kwikset Corp. v. Super. Ct.*, 51 Cal. 4th 310, 324 (2011). A private plaintiff may only sue under the UCL if it has "suffered injury in fact and has lost money or property." Cal. Bus. & Prof. Code § 17204. Standing for the UCL is "substantially narrower than federal standing under Article III." *Kwikset*, 51 Cal. 4th at 324; *see also McGee*, 982 F.3d at 705 n.3 (same). Here, because Plaintiffs lack Article III standing, *see* Part I, *supra*, they certainly lack UCL standing. Plaintiffs do not plausibly allege the required "lost money or property." Cal. Bus. & Prof. Code § 17204. They assert that "Plaintiffs and the Class have suffered monetary damages as a result of GitHub's and OpenAI's conduct." Compl. ¶ 213. But Plaintiffs do not allege any factual theory of *how* they suffered monetary loss as a result of the training of Copilot or its suggestions, let alone any specific allegations supporting such a theory. Their "[t]hreadbare recital[] of the elements of [the] cause of action, supported by mere conclusory statements, do[es] not suffice." *Iqbal*, 556 U.S. at 678.

In all events, Plaintiffs' UCL claim fails because it is dependent upon otherwise defective DMCA, tort, passing off, and privacy violations. "[T]o state a claim under the unlawful prong of the UCL, a plaintiff must sufficiently plead a predicate violation." *Lopez v. Bank of Am., N.A.*, 505 F. Supp. 3d 961, 976 (N.D. Cal. 2020) (quotation marks omitted). Plaintiffs did not.

V. PLAINTIFFS' REVERSE PASSING OFF CLAIM IS BARRED UNDER *DASTAR*.

Plaintiffs purport to assert a claim under the Lanham Act for "reverse passing off." Compl. at 43 (Count V). The allegation that Defendants have marketed, distributed, or sold works in which Plaintiffs claim some interest "without attribution" is a copyright claim masquerading as a trademark claim, and therefore squarely barred by the Supreme Court's decision in *Dastar Corp. v. Twentieth Century Fox Film Corp.*, 539 U.S. 23, 37 (2003); *see also Sybersound Records, Inc. v. UAV Corp.*, 517 F.3d 1137, 1143-44 (9th Cir. 2008).

VI. PLAINTIFFS' PRIVACY CLAIMS AGAINST GITHUB FAIL.

Plaintiffs attempt to claim that GitHub invaded their privacy (Counts VIII, IX, and X) at the same time that the overall thrust of the Complaint is a failure to publicly attribute their identity to code. The Complaint does not identify any specific privacy-related provision of any alleged contract that was violated or how; does not allege a data breach that could yield liability under the California Consumer Privacy Act ("CCPA"); and does not identify any personal data that was negligently handled nor how that occurred. As noted in Part I.B, *supra*, no facts in the Complaint explain how Plaintiffs were harmed by any privacy-based conduct by GitHub.

Whatever privacy-based claim Plaintiffs purport to advance, at a minimum they must provide a description of the PII at issue and a plausible allegation of some conduct in connection with that PII. Plaintiffs fail to provide those basics. *See* Compl. ¶¶ 215-39. The Complaint uses the conclusory terms "personal data," "personal information," and "PII" in describing Plaintiffs' claims, *e.g.*, Compl. ¶¶ 220, 228-29, but never defines those terms or specifies the actual data types discussed. Similarly, the Complaint insinuates that GitHub is collecting, distributing, using, or selling such information, but does so without alleging any facts. *See* Compl. ¶¶ 220, 229-33, 236-37. Plaintiffs never identify which personal information is involved, how it is captured, shared, sold, or otherwise distributed, or how any of this has resulted or is likely to result in concrete injury. Plaintiffs thus fail to provide basic notice of their claims for breach of GitHub's Privacy Policy, violation of the CCPA, or negligence, necessitating dismissal.

The CCPA claim, moreover, faces a particularly demanding standard that Plaintiffs do not even attempt to satisfy. The CCPA provides a limited private right of action. *See* Cal. Civ. Code § 1798.150. It allows a private claim only if a consumer's "nonencrypted or nonredacted personal information, as defined in [§ 1798.81.5(d)(1)(A)], is subject to an unauthorized access and exfiltration, theft, or disclosure as a result of the business's violation of the duty to implement and maintain reasonable security procedures and practices appropriate to the nature of the information to protect the personal information." It is thus § 1798.81.5(d)(1)(A)—not § 1798.140(o)(1), as Plaintiffs allege, Compl. ¶ 228—that defines personal information for purposes of a CCPA claim. That provision defines such information as last name and first name or first initial, *plus* one or more of a social security number, driver's license or similar number, account number or card number with access code or password that would give access to a financial account, medical information, health insurance information, biometric data, or genetic data. *See* Cal. Civ. Code § 1798.81.5(d). Plaintiffs' allegations of misuse of unspecified PII are not actionable under § 1798.150, because Plaintiffs fail to plead any facts regarding data security or theft of the specified categories of data that could support a CCPA claim. Simply parroting the language of the statute is not enough to plausibly plead a claim, *see Cork v. CC-Palo Alto, Inc.*, 534 F. Supp. 3d 1156, 1183 (N.D. Cal. 2021), and that is particularly true here where the facts described in the Complaint regarding training a machine learning model on publicly available data appears to be far afield from the ordinary fact patterns in data breach cases.

VII. PLAINTIFFS' UNJUST ENRICHMENT, CONSPIRACY, AND DECLARATORY RELIEF ALLEGATIONS FAIL ALONG WITH THE UNDERLYING CLAIMS.

Plaintiffs purport to advance an unjust enrichment claim against GitHub (Count VI), but "there is no cause of action in California for unjust enrichment." *Melchior v. New Line Prods., Inc.*, 106 Cal. App. 4th 779, 793 (2003); *Bosinger v. Belden CDT, Inc.*, 358 F. App'x 812, 815 (9th Cir. 2009). Nor is civil conspiracy (Count XI) a cause of action under California law. *Applied Equip. Corp. v. Litton Saudi Arabia Ltd.*, 7 Cal. 4th 503, 514 (1994).

Plaintiffs' declaratory relief claim (Count XII) rests on the viability of underlying claims, and should be dismissed along with them. Count XII also runs afoul of Article III because the Complaint does not set forth a specific requested declaration, let alone make allegations demonstrating an injury that would be redressed by a declaration. *See* Part I, *supra*; *Lujan*, 504 U.S. at 561; *see Hobbs v. Sprague*, 87 F. Supp. 2d 1007, 1012 (N.D. Cal. 2000) (no standing where Complaint fails to show that an alleged injury would be redressed by a favorable decision).

CONCLUSION

The Court should grant the motion to dismiss the Complaint in its entirety.

Dated: January 26, 2023

Orrick, Herrington & Sutcliffe LLP

By: */s/ Annette L. Hurst*
ANNETTE L. HURST
Attorneys for Defendants
GitHub, Inc. and Microsoft Corp.