

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

United States District Court
Northern District of California

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

J. DOE 1, et al.,
Plaintiffs,
v.
GITHUB, INC., et al.,
Defendants.

Case No. 22-cv-06823-JST

**ORDER GRANTING IN PART AND
DENYING IN PART MOTIONS TO
DISMISS**

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 licenses³ 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 _____
26 ³ These eleven licenses are (1) Apache License 2.0; (2) GNU General Public License version 3
27 (“GPL-3.0”); (3) MIT License; (4) The 2-Clause BSD License (“BSD 2”); (5) The 3-Clause BSD
28 License (“BSD 3”); (6) Boost Software License; (7) Eclipse Public License 2.0; (8) GNU Affero
29 General Public License version 3 (“AGPL-3.0”); GNU General Public License version 2 (“GPL
30 2”); (10) GNU Lesser General Public License version 2.1 (“LGPL-2.1”); and (11) Mozilla Public
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 _____
27 ⁴ The “Injunctive Relief Class” and “Damages Class” are each defined as: “All persons or entities
28 domiciled in the United States that[] (1) owned an interest in at least one [U.S.] 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 between January 1, 2015 and the present
(the ‘Class Period’).” Compl. ¶ 34 (footnote omitted).

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
 24 applies to each claim. “[D]ue to variances among state laws, failure to allege which state law
 25 governs a common law claim is grounds for dismissal.” *In re Nexus 6P Prods. Liab. Litig.*, 293 F.
 26 Supp. 3d 888, 933 (N.D. Cal. 2018) (quoting *Romero v. Flowers Bakeries, LLC*, No. 14-cv-05189-
 27 BLF, 2016 WL 469370, at *12 (N.D. Cal. Feb. 8, 2016)). With respect to the state law claims in
 28 any future amended complaint, Plaintiffs shall identify the state under whose law the claim is
 29 brought. When claims which share a legal theory are brought under multiple laws (for example, a
 30 federal statute, a state statute, and common law, such as Plaintiffs’ claim for unfair competition),
 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
 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 *Mendondo 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

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.