EXHIBIT A

FOR PUBLICATION

UNITED STATES COURT OF APPEALS
FOR THE NINTH CIRCUIT

J. DOE, 1; J. DOE, 2-5, individually and on behalf of all others similarly situated,

Plaintiffs - Appellants,

v.

GITHUB, INC., a Delaware corporation; MICROSOFT CORPORATION, a Washington corporation; OPENAI, INC., a Delaware nonprofit corporation; OPENAI, LP, a Delaware limited partnership; OPENAI GP, LLC, a Delaware limited liability company; OPENAI STARTUP FUND GP I, LLC, a Delaware limited liability company; OPENAI STARTUP FUND I, LP, a Delaware limited partnership; OPENAI STARTUP FUND MANAGEMENT, LLC, a Delaware limited liability company; OPENAI OPCO, LLC; OPENAI, LLC; OPENAI GLOBAL, LLC; OAI CORPORATION; OPENAI HOLDINGS, LLC; OPENAI

No. 24-7700

D.C. No. 4:22-cv-06823-JST

OPINION

HOLDCO, LLC; OPENAI INVESTMENT; OPENAI STARTUP FUND SPV I, LP; OPENAI STARTUP FUND SPV GP I, LLC,

Defendants - Appellees.

Appeal from the United States District Court
for the Northern District of California
Jon S. Tigar, District Judge, Presiding

Argued and Submitted February 11, 2026
San Francisco, California

Filed September 16, 2026

Before: Sidney R. Thomas and Eric D. Miller, Circuit Judges, and Stanley Blumenfeld, Jr., District Judge.*

Opinion by Judge Miller

* The Honorable Stanley Blumenfeld, Jr., United States District Judge for the Central District of California, sitting by designation.

SUMMARY**

Digital Millennium Copyright Act

In an interlocutory appeal pursuant to 28 U.S.C. § 1292(b), the panel affirmed the district court's dismissal of part of an action brought under the Digital Millennium Copyright Act ("DMCA") by programmers against GitHub, Inc., and other defendants.

GitHub is a platform that allows developers to store, manage, and share their code. GitHub Copilot and Codex are tools that help programmers write code using artificial intelligence—specifically, large language models that were trained on millions of software projects on GitHub. Plaintiffs alleged that Copilot and Codex reproduced their code without attribution in violation of 18 U.S.C. § 1202(b), a provision of the DMCA.

The panel concluded that plaintiffs had Article III standing to bring their DMCA claim because they plausibly alleged a substantial risk of injury.

The panel nonetheless held that plaintiffs did not state a claim under the DMCA because their allegations under an "output" theory showed that Copilot and Codex do not "remove or alter" copyright management information from a copy of an existing protected work but instead create new works that never contained that information.

** This summary constitutes no part of the opinion of the court. It has been prepared by court staff for the convenience of the reader.

The panel declined to consider, as forfeited, plaintiffs' "input" theory of DMCA liability.

COUNSEL

Jesse Panuccio (argued), Daniel Morales, and Evan M. Ezray, Boies Schiller Flexner LLP, Fort Lauderdale, Florida; Joshua M. Stein, Maxwell V. Pritt, and Margaux Poueymirou, Boies Schiller Flexner LLP, San Francisco, California; Matthew Butterick, Butterick Law PC, Los Angeles, California; Joseph R. Saveri, Christopher K.L. Young, William W.C. Guardado, and Evan Creutz, Saveri Law Firm LLP, San Francisco, California; for Plaintiffs-Appellants.

Lisa S. Blatt (argued), Erin M. Sielaff, Claire R. Cahill, Thomas G. Hentoff, and Claire L. Lazar, Williams & Connolly LLP, Washington, D.C.; Christopher J. Cariello (argued), Orrick Herrington & Sutcliffe LLP, New York, New York; Alyssa M. Caridis, and Geoffrey C. Shaw, Orrick Herrington & Sutcliffe LLP, Los Angeles, California; Jonas Q. Wang, Orrick Herrington & Sutcliffe LLP, Washington, D.C.; Nicholas Gonzalez and Annette L. Hurst, Orrick Herrington & Sutcliffe LLP, San Francisco, California; Aileen McGrath, Joseph C. Gratz, and Tiffany Cheung, Morrison & Foerster LLP, San Francisco, California; John R. Lanham, Morrison & Foerster LLP, San Diego, California; for Defendants-Appellees.

Brian Scarpelli and Priya Nair, ACT The App Association, Washington, D.C., for Amicus Curiae ACT The App Association.

Paul N. Harold, Wilson Sonsini Goodrich & Rosati PC, Washington, D.C., for Amici Curiae Chamber of Progress and Computer & Communications Industry Association.

Robert W. Clarida, Reitler Kailas & Rosenblatt LLP, New York, New York, for Amici Curiae The Authors Guild Inc., The Association of American Publishers, The News/Media Alliance, and The International Association of Scientific Technical and Medical Publishers.

Corynne McSherry and Victoria Noble, Electronic Frontier Foundation, San Francisco, California, for Amici Curiae Electronic Frontier Foundation and Public Knowledge.

Christopher Bavitz, Cyberlaw Clinic, Harvard Law School, Cambridge, Massachusetts, for Amicus Curiae Authors Alliance.

Jennifer M. Urban and Erik Stallman, Samuelson Law Technology & Public Policy Clinic, University of California at Berkeley School of Law, Berkeley, California, for Amici Curiae Intellectual Property Law Professors.

OPINION

MILLER, Circuit Judge:

GitHub is a platform that allows developers to store, manage, and share their code. GitHub Copilot (Copilot) and Codex are tools that help programmers write code using artificial intelligence—specifically, large language models that were trained on millions of software projects on GitHub. Plaintiffs are programmers who published open-source code on GitHub. They believe that Copilot and Codex may

reproduce their code without attribution and allege that, were that to happen, it would violate 17 U.S.C. § 1202(b), enacted as part of the Digital Millennium Copyright Act (DMCA), Pub. L. No. 105-304, § 103, 112 Stat. 2860, 2872 (1998). Plaintiffs brought this action under the DMCA, and the district court dismissed plaintiffs' DMCA claims. We agree with the district court that plaintiffs do not state a claim under the DMCA because their allegations show that Copilot and Codex do not "remove or alter" copyright management information (CMI) from a copy of an existing protected work but instead create new works that never contained that information. *See* 17 U.S.C. § 1202(b)(1). We therefore affirm.

I

Microsoft owns and operates GitHub, which is the world's largest hosting service for open-source software—that is, software whose creators choose to make its source code publicly available. Much of the code in public repositories on GitHub is made available under open-source licenses that permit other developers to use or modify that code subject to certain conditions. One of the most common conditions is attribution, meaning that a copy of the license—including the author's name and copyright notice—must be included with any copy or derivative of the code.

Microsoft is also a major investor in OpenAI, a company that develops and sells artificial intelligence models. One of OpenAI's projects is Codex, a generative AI tool that works autonomously to draft code based on a programmer's prompts. Copilot is a paid subscription service jointly developed by GitHub and OpenAI that uses a modified version of Codex to produce blocks of code in response to a

user's prompts. In all respects relevant to this case, Codex and Copilot function similarly, so for simplicity, we will refer solely to Copilot.

As described in the complaint, Copilot is a large language model that works by identifying statistical patterns in the structure of code. The model relies on those patterns to identify the "most statistically likely completion, based on the examples it reviewed in training," to a particular prompt or snippet of code. To identify relevant statistical patterns, Copilot was trained on billions of lines of publicly available code, including code from "all available public GitHub repositories."

Plaintiffs are programmers who have published copyrighted code, subject to various licenses, on public GitHub repositories. They brought this putative class action against GitHub, Microsoft, and OpenAI in the Northern District of California. After two rounds of dismissals and amendments, the complaint has been whittled down to three claims: one claim under the DMCA and two claims for breach of contract.

Plaintiffs' DMCA claim rests on their allegation that Copilot will sometimes produce "identical copies of code Copilot was trained on," a possibility that "is only increasing as the model scales." The operative complaint includes examples of cases in which Copilot reproduced "essentially verbatim" copies of code written by the proposed class representatives. Plaintiffs allege that when Copilot reproduces code from its training materials, it does not include the attribution, copyright notices, or other license terms that were included with the code when it was published on GitHub. Plaintiffs claim that this omission violates section 1202(b) of the DMCA.

"In enacting the DMCA, Congress sought to mitigate the problems presented by copyright enforcement in the digital age." *MDY Indus., LLC v. Blizzard Ent., Inc.*, 629 F.3d 928, 942 (9th Cir. 2010) (citing *Universal City Studios, Inc. v. Corley*, 273 F.3d 429, 440 (2d Cir. 2001)). As relevant here, section 1202 makes it unlawful to (1) "intentionally remove or alter any [CMI]," (2) "distribute or import for distribution [CMI] knowing that the [CMI] has been removed or altered without authority of the copyright owner or the law," or (3) "distribute, import for distribution, or publicly perform works, copies of works, or phonorecords, knowing that [CMI] has been removed or altered without authority of the copyright owner or the law." 17 U.S.C. § 1202(b). The DMCA defines CMI as certain "information conveyed in connection with copies . . . of a work," including "[t]he title and other information identifying the work, including the information set forth on a notice of copyright," "[t]he name of, and other identifying information about, the author" or "the copyright owner of the work," and the "[t]erms and conditions for use of the work." *Id.* § 1202(c). The DMCA does not require works to include CMI or prescribe the particular types of CMI that may be used. But if a work does contain CMI, section 1202(b) protects that CMI from deliberate removal or alteration.

The district court dismissed the DMCA claim under Federal Rule of Civil Procedure 12(b)(6), first with leave to amend, and then with prejudice. The court reasoned "that Section 1202(b) claims require that copies be 'identical.'" It emphasized plaintiffs' allegations that Copilot's output is often a "near-identical copy" or "modified copy" of code on GitHub, explaining that the complaint's examples all involved instances in which "the Copilot output is a 'modified format,' 'variation[],' or the 'functional[]

Case: 24-7700, 09/16/2026, DktEntry: 126.1, Page 10 of 128) Page 1

equivalent' of the licensed code." The court concluded that those allegations were not sufficient to state a claim under the DMCA because plaintiffs' code and Copilot's output "were not identical." The court denied a motion to dismiss the contract claims, which remain pending.

The district court certified its order dismissing the section 1202(b) claims for interlocutory appeal under 28 U.S.C. § 1292(b). The court held that "whether Sections 1202(b)(1) and (b)(3) of the DMCA impose an identicality requirement" is a controlling question of law on which there is substantial ground for difference of opinion.

We review de novo the district court's grant of a motion to dismiss. *Wells Fargo Bank, N.A. v. Mahogany Meadows Ave. Tr.*, 979 F.3d 1209, 1213 (9th Cir. 2020). "To survive a motion to dismiss, a complaint must contain sufficient factual matter, accepted as true, to 'state a claim to relief that is plausible on its face.'" *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) (quoting *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570 (2007)).

II

Plaintiffs have articulated two different theories of DMCA violations. The first, which we will call the "input" theory, is that defendants violated section 1202(b)(1) at the training stage by removing CMI from putative class members' code before feeding the stripped code into Copilot as training data. The second, which we will call the "output" theory, is that Copilot sometimes returns memorized training data as output to users without including the source code's CMI, and that in doing so Copilot "remove[s]" or "alter[s]" CMI on copies of plaintiffs' protected work, in violation of section 1202(b)(1) and (3).

We decline to consider the input theory because plaintiffs forfeited it. Read in isolation, the complaint might be understood to assert such a theory. For example, the complaint alleges that "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." But although plaintiffs at times gestured toward an input theory, the district court made clear that it did not understand plaintiffs to have asserted a claim based on that theory. At the hearing on defendants' first motion to dismiss, the court asked whether the act of copying training data into Copilot violated the attribution requirement of open-source licenses, and plaintiffs' counsel answered, "Perhaps it doesn't." The court then observed that the "complaint is not about training. It just isn't." And in the district court's subsequent order granting in part defendants' motion to dismiss plaintiffs' first complaint, the court stated unequivocally: "Plaintiffs do not allege they were injured by Defendants' use of licensed code as training data."

If plaintiffs believed that the district court misunderstood their position, it was incumbent upon them to say so. But plaintiffs did not do so, either orally or in writing. In arguing that they preserved an input theory, plaintiffs point only to their statement that "[t]he DMCA makes the mere removal of CMI from digital copies illegal before distribution of copies" and to one of OpenAI's filings, in which OpenAI purportedly acknowledged that plaintiffs advanced a claim based on "the pre-distribution removal of CMI." Plaintiffs' statement was part of their explanation of the distinction between section 1202(b)(1) and section 1202(b)(3) and did not suggest disagreement with the district court's conclusion that the complaint did not advance an input theory. And the

cited OpenAI filing argued that plaintiffs' second amended complaint "admits that no pre-distribution [CMI] removal occurred" and does not allege any claim based on a pre-distribution removal theory. Plaintiffs have identified nothing in their district court briefing after the first motion to dismiss that would have put the court on notice that they wished to pursue an input theory. We therefore conclude that plaintiffs forfeited the theory.

III

As to the output theory, defendants argue that plaintiffs lack standing because they did not adequately allege that their code will be subject to copying by Copilot without reproduction of the accompanying CMI. The district court did not address standing in its order granting the motion to dismiss, and the question certified for appeal does not discuss standing. But our jurisdiction under section 1292(b) "applies to the *order* certified to the court of appeals, and is not tied to the particular question formulated by the district court." *Yamaha Motor Corp., U.S.A. v. Calhoun*, 516 U.S. 199, 205 (1996). We therefore "may address any issue fairly included within the certified order." *Id.* That includes the question of standing. Although that question was "not expressly certified," the district court had addressed it in an earlier order, and it is "fairly raised by the order under review." *Lee v. American Nat'l Ins. Co.*, 260 F.3d 997, 1001 (9th Cir. 2001).

To establish Article III standing, a plaintiff must show, among other things, that it faces an injury in fact—that is, "an invasion of a legally protected interest which is (a) concrete and particularized and (b) actual or imminent, not conjectural or hypothetical." *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560 (1992) (citation, internal

quotation marks, and footnote omitted). "An allegation of future injury" can establish standing if "there is a substantial risk that the harm will occur." *Susan B. Anthony List v. Driehaus*, 573 U.S. 149, 158 (2014) (citation and internal quotation marks omitted). Allegations of standing "must be supported in the same way as any other matter on which the plaintiff bears the burden of proof, *i.e.*, with the manner and degree of evidence required at the successive stages of the litigation." *Lujan*, 504 U.S. at 561. "At the pleading stage, general factual allegations of injury resulting from the defendant's conduct may suffice" because courts must "presum[e] that general allegations embrace those specific facts that are necessary to support the claim." *Id.* (quoting *Lujan v. National Wildlife Fed'n*, 497 U.S. 871, 889 (1990)).

According to defendants, even if Copilot makes copies of code without CMI, the risk that it will reproduce plaintiffs' code—rather than the code of some other GitHub contributor—is "no more than conjecture." *City of Los Angeles v. Lyons*, 461 U.S. 95, 108 (1983). We disagree. The complaint alleges a substantial risk of injury because it states that "Copilot reproduces the code of the Named Plaintiffs without attribution," "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," that "[i]t is therefore likely if not certain that verbatim, near-verbatim or modified copies of each Plaintiff['s] code have already been emitted by Copilot," and that "each day these violations must be accruing with astonishing frequency."

Of course, a complaint must "contain sufficient factual matter" to make the stated allegations plausible. *Iqbal*, 556 U.S. at 678. But the complaint here meets that standard. The

complaint cites academic research finding that large language models will sometimes "emit the memorized training data verbatim," a phenomenon that "will likely get worse as models continue[] to scale." The complaint notes that GitHub has implemented a duplicate-detection feature that allows users to block output suggestions matching public code—specifically, verbatim snippets of 150 characters or more—which is some evidence that Copilot can and does emit literally identical copies of code. And the complaint also provides examples of cases in which Copilot reproduced portions of the named plaintiffs' code verbatim.

If this case were to proceed to summary judgment, plaintiffs would need to present evidence sufficient to create a genuine factual issue as to whether they face a "substantial risk" of harm. *Susan B. Anthony List*, 573 U.S. at 158. We need not determine whether the evidence referred to in the complaint would be sufficient to meet that standard. But the factual allegations in the complaint make it plausible that plaintiffs face a substantial risk, and that is enough to survive a motion to dismiss.

IV

Turning to the merits of the output theory, we agree with the district court that the complaint does not state a claim under the DMCA. The complaint describes an AI tool that does not look up and reproduce stored work but rather creates new work. That new work may or may not infringe plaintiffs' copyrights in their code, but it cannot reasonably be described as a copy of that code from which CMI has been "removed" or "altered."

"[W]e start where we always do: with the text of the statute." *Van Buren v. United States*, 593 U.S. 374, 381 (2021). Section 1202(b) provides that "[n]o person shall,

without the authority of the copyright owner or the law," "intentionally remove or alter any copyright management information" or "distribute . . . works [or] copies of works, . . . knowing that copyright management information has been removed or altered without authority of the copyright owner or the law." 17 U.S.C. § 1202(b)(1), (3). The verb "remove" means "to get rid of," while "alter" means "to cause to become different in some particular characteristic." *Webster's Third New International Dictionary* 63, 1921 (1993). In context, both verbs imply taking an affirmative act with respect to CMI connected to a work that already exists. *See, e.g.*, 4 Nimmer on Copyright § 12A.10[B][1][a] (2026) (citing "defacing . . . the title page of a book" or "delet[ing] the electronic information that may accompany a computer file" as paradigmatic section 1202(b) violations). One who creates a new work and fails to include CMI cannot be said to have "removed" or "altered" anything. *See Falkner v. General Motors LLC*, 393 F. Supp. 3d 927, 938 (C.D. Cal. 2018) (explaining that mere "failure to include" CMI does not "constitute[] removal or alteration" of CMI); *accord Kipp Flores Architects, LLC v. AMH Creekside Dev., LLC*, No. 23-50750, 2026 WL 2453096, at *13–17 (5th Cir. Aug. 21, 2026) (opinion of Willett, J.). Instead, removal or alteration requires taking away or changing attribution information from an existing work.

The statutory definition of CMI reinforces this reading. CMI is "information conveyed in connection with copies . . . *of a work*," not in connection with excerpts or derivative works. 17 U.S.C. § 1202(c) (emphasis added). A "copy" of a work is a "material object[] . . . in which a work is fixed," *id.* § 101, so CMI is information that exists in connection with those material objects. Violating section 1202(b) requires removing or altering the CMI on those material

objects, such as by reprinting an image while cropping out the "printed gutter credit near the [i]mage" identifying the photographer. *Murphy v. Millennium Radio Grp.*, 650 F.3d 295, 305 (3d Cir. 2011); *see also Falkner*, 393 F. Supp. 3d at 938–39 (photographer did not remove or alter CMI when he photographed a mural from an angle that rendered the artist's signature not visible). To state a claim under section 1202, plaintiffs must therefore allege that defendants removed or altered CMI from copies of existing protected works; merely alleging that a similar or derivative work does not include the CMI, without facts showing its removal or alteration, is insufficient.

The district court and defendants characterize this principle as the DMCA's "identicality" requirement. But "identicality" is something of a misnomer because the DMCA does not require literal identicality between the plaintiff's work and the allegedly infringing work. The concept of "identicality" is best understood as a gloss on the statutory terms "remove," "alter," and "copies" rather than an independent (and atextual) element of a section 1202(b) claim.

To the extent identicality is relevant, it is because one way to determine whether CMI has been "remove[d]" from a copyrighted work is to compare the allegedly infringing copy with the original. If the two works are otherwise identical—sharing, for example, the same composition, cropping, or other distinctive features—but the allegedly infringing version omits the CMI that accompanied the original, a factfinder may reasonably infer that the defendant removed CMI from the copied work. We relied on just such a comparison in *Friedman v. Live Nation Merchandise, Inc.*, where we held that because the allegedly infringing photographs were "exact copies" of the images as they

appeared in the copyright owner's book, and "[b]ecause the *only* material difference" in the allegedly infringing versions "was that the CMI was missing, it was necessarily the case that the CMI had been removed on the copied version." 833 F.3d 1180, 1188 (9th Cir. 2016). Although substantial reproduction of a plaintiff's existing work without CMI is not always dispositive, it will often be strong circumstantial evidence that the defendant removed CMI from a copy of the plaintiff's work. On the other hand, material differences between the works may suggest that the defendant instead created a new, derivative work to which the plaintiff's CMI was never attached and from which the defendant could not have "removed" or "altered" CMI.

Plaintiffs observe that requiring literal identicality would make it easy to circumvent the DMCA: So long as the defendant changed one word on one page of a book, in addition to deleting the CMI, that person would avoid violating section 1202(b) because the original and allegedly infringing versions would not be identical. But as defendants concede, the two works need not be literally identical to support an inference that CMI has been removed. *See Real World Media LLC v. Daily Caller, Inc.*, 744 F. Supp. 3d 24, 40 (D.D.C. 2024) ("[I]t would be odd if a defendant could evade DMCA liability by removing or altering CMI in a copied work but only disseminating 99% rather than 100% of that work."). Minor cosmetic changes will not necessarily protect a defendant who substantially or entirely reproduces the protected work and removes CMI. *See New York Times Co. v. Microsoft Corp.*, 777 F. Supp. 3d 283, 317 (S.D.N.Y. 2025).

But that concession does not help plaintiffs because they have not alleged removal of CMI from a copy of their work, even one with minor alterations. In reaching that conclusion,

we recognize the limitation of relying on examples involving traditional print media, such as when a defendant has "defac[ed] . . . the title page of a book," 4 Nimmer on Copyright § 12A.10[B][1][a] (2026). While such examples may provide a useful starting point for reasoning by analogy, they do not map neatly onto the emerging digital technologies like artificial intelligence to which the DMCA's protections also apply. Even so, plaintiffs' own allegations about Copilot show that it is best understood as learning from existing works and then creating new works based on that learning process, not as making copies of existing works.

According to the complaint, Copilot relies on a "complex probabilistic process" to predict "the most likely solution to a given prompt," based on "the solution it has found in the most projects" answering similar questions. Sometimes, plaintiffs allege, that output may match snippets of "code from the training data." But their account of the algorithm's internal process—inferring "statistical patterns governing the structure of code" and identifying the most likely completion—does not describe an action taken with respect to CMI attached to an existing work. Instead, it describes a process through which Copilot generates new works. In that respect, it differs from a traditional search engine, which, in response to a user's query, retrieves and displays stored information—that is, copies of materials that already exist. If Copilot functioned like a search engine and produced outputs that were identical to plaintiffs' code but did not contain CMI, then plaintiffs might have a stronger claim that Copilot had removed their CMI. But that is not what plaintiffs have alleged.

To be sure, Copilot's output may in some cases be substantially similar to existing code. As the district court

observed, Copilot may produce "a 'modified format,' 'variation[],' or the 'functional[] equivalent' of the licensed code" without the CMI. We express no view on whether that similarity would allow plaintiffs to assert a claim for copyright infringement. But we note that many copyright cases involve the creation of a work that is substantially similar to the plaintiff's without attribution. *See Yonay v. Paramount Pictures Corp.*, 163 F.4th 685, 692 (9th Cir. 2026) ("To show unlawful appropriation," a plaintiff must show "that the works in question share 'substantial similarity in protectable expression.'" (emphasis omitted) (quoting *Skidmore as Tr. for Randy Craig Wolfe Tr. v. Led Zeppelin*, 952 F.3d 1051, 1064 (9th Cir. 2020) (en banc))). If that were all it took to violate section 1202(b), the DMCA would supplant traditional copyright protections and subject defendants to potentially ruinous liability under the DMCA's enhanced statutory damages. *Compare* 17 U.S.C. § 1203(c)(3) (permitting up to $25,000 *per violation*) *with* 17 U.S.C. § 504(c)(1) (capping traditional copyright statutory damages at $30,000 *per work*). We decline plaintiffs' invitation to transform run-of-the-mill copyright-infringement claims into DMCA claims.

AFFIRMED.