

Journalists' Lack of Harm Fatal to DMCA Claims against AI Developer

November 12, 2024

Developers of artificial intelligence (“AI”) systems notched a victory last week when a federal judge dismissed claims under the Digital Millennium Copyright Act (“DMCA”) premised on the use of copyrighted works in AI training data, holding that the plaintiffs had failed to show any concrete harm and therefore lacked standing to bring their claims. *Raw Story Media, Inc. v. OpenAI Inc.*, No. 24 CV 01514-CM, 2024 WL 4711729 (S.D.N.Y. Nov. 7, 2024).

The plaintiffs, two news organizations, alleged that OpenAI had used their copyrighted works in training ChatGPT, one of the most prominent generative AI “chatbots” on the market today. The journalists did not assert straightforward copyright infringement claims: they instead argued that OpenAI had removed copyright management information (“CMI”) before using their works to train ChatGPT, in violation of the DMCA.

Judge Colleen McMahon of the Southern District of New York dismissed plaintiffs’ suit in its entirety, holding that plaintiffs had no cognizable claim for damages or injunctive relief because they failed at this stage of litigation to demonstrate that they had been harmed in any way by OpenAI’s actions. The key flaw in plaintiffs’ case, according to Judge McMahon, was the absence of any evidence showing that ChatGPT had in fact disseminated (or was even likely to disseminate) plaintiffs’ copyrighted work without their CMI present.

Below, we provide a background on the DMCA’s provisions governing CMI—known as Section 1202(b)—and recent DMCA litigation against AI developers. The recent decisions, including the *Raw Story Media* holding, make clear that up until this point plaintiffs have faced an uphill battle in seeking to hold generative AI platforms liable under the DMCA.

Section 1202’s Protections for CMI—and Their Limits. The advent of widespread personal computing and the Internet in the 1990s made it easier than ever to distribute content—and to make unauthorized copies of that content. In response to widespread concern among rights holders about the ease with which consumers could make

inexpensive copies of music and movies, in 1998, Congress enacted the DMCA to bring copyright law into the Internet age. Generally, the DMCA provided additional digital rights management (“DRM”) and copyright protections to aid rights holders in protecting their intellectual property assets by prohibiting the production and distribution of technology that attempts to circumvent DRM and the act of such circumvention. Among the DMCA’s provisions were Section 1202’s protections for CMI: identifying information about the source of the copyrighted work and its owner that is commonly attached to the work via a physical marking, such as a watermark, or in the metadata of file, and which also includes information like the name of an author on an online article. See 17 U.S.C. § 1202.

When copyright owners include CMI alongside their copyrighted work, Section 1202(b) prohibits altering or removing that CMI or distributing copies of the copyrighted work on which the CMI has been altered or removed. To plead a violation of Section 1202(b), a copyright holder must also establish that the defendant knew, or had reasonable grounds to know, that their actions would “induce, enable, facilitate, or conceal” copyright infringement.

While the statutory elements of a Section 1202(b) claim appear straightforward, certain courts considering claims involving CMI have added two hurdles for copyright plaintiffs: a “double scienter” requirement and an identity requirement.

Double Scienter. In the Second, Ninth, and Eleventh Circuits, a plaintiff bringing a claim under Section 1202(b) must demonstrate that the infringer had knowledge *both* that CMI was altered or removed *and* that the alteration or removal of CMI would likely result in copyright infringement. See, e.g., *Mango v. Buzzfeed*, 970 F.3d 167, 171 (2Cir. 2020); *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 675 (9th Cir. 2018); *Victor Elias Photography, LLC v. Ice Portal, Inc.*, 43 F.4th 1313, 1320 (11th Cir. 2022), *cert. denied*, 143 S. Ct. 736, 214 L. Ed. 2d 385 (2023).

Identity. Some courts considering Section 1202 claims have also required that the work at issue be reproduced *exactly*. In other words, the plaintiff must show that (aside from the removal or alteration of their CMI), the defendants made an exact copy of the original copyrighted work. See, e.g., *Kirk Kara Corp. v. Western Stone and Metal Corp.*, No. CV 2-20-1931-DMG, 2020 WL 5991503, at *6 (C.D. Cal. 2020) (no DMCA Section 1202(b) violation because the work with the removed CMI, while “substantially similar,” was not an “identical” copy of the plaintiff’s work); *Frost-Tsuji Architects v. Highway Inn, Inc.*, No. 13-00496 SOM/BMK, 2015 WL 263556, at *3 (D. Haw. Jan. 21, 2015), *aff’d*, 700 F. App’x 674 (9th Cir. 2017) (no Section 1202(b) violation where an allegedly infringing drawing was “not identical” to the copyrighted work).

AI Developers' Recent Successes in Dismissing Section 1202 Claims. In *Raw Story Media*, the plaintiffs were journalists who published news and other articles on their websites. While plaintiffs were unable to allege specific facts about the training data behind the current version of ChatGPT because that information is not public, they alleged that earlier versions were trained using WebText, WebText2 and Common Crawl, which are large open repositories of data scraped from the internet. Plaintiffs alleged that their copyright-protected works were posted online with CMI and that copies of their work were in the data repositories used to train defendants' generative AI program—but that defendants intentionally removed that CMI before it was used.

Plaintiffs could not allege that ChatGPT had actually reproduced or disseminated their copyrighted work without their CMI, which is the core of a Section 1202 claim. The court held that the mere use of the copyrighted work in the training data was not the kind of harm that Section 1202 was designed to prohibit because it did not involve reproduction of the copyrighted material without the CMI attached. The court did, however, give plaintiffs a chance to plead to attempt to establish standing or pursue a different legal theory.

The *Raw Story Media* case was not the first time AI developers have succeeded in avoiding liability under Section 1202 for their training data. Other plaintiffs' DMCA claims have similarly been dismissed for failure to allege the necessary violations in ways that show how the difficult hurdles of both double scienter and identity have proven to be stumbling blocks for claims involving generative AI.

In 2023, an author's Section 1202(b) claim against Meta over training data for its AI model known as LLaMA was dismissed for failure to show that LLaMA was actually generating and distributing copies of plaintiff's work, "much less" that such distribution was done with false or altered CMI. *Kadrey v. Meta Platforms, Inc.*, No. 23 CV 03417-VC, 2023 WL 8039640, at *2 (N.D. Cal. Nov. 20, 2023). Without alleging any facts to support those elements, the defendants' DMCA claim was summarily dismissed.

A different plaintiff seeking to hold GitHub, Microsoft and OpenAI liable under Section 1202(b) was dealt a similar defeat on their DMCA claims. Although the court initially agreed that the plaintiff had pled facts sufficient to support an allegation that defendants had intentionally removed CMI and doing so carried a risk of copyright infringement, the court later held that the plaintiffs' DMCA claims nevertheless failed because allegations that an AI system produced *modified* versions of the plaintiff's work did not state a DMCA claim. *Doe 1 v. GitHub, Inc.*, No. 22 CV 6823, 2024 WL 235217, at *9 (N.D. Cal. Jan. 22, 2024).

Compared to these earlier cases, *Raw Story Media* is notable because the DMCA Section 1202(b) violation was the *only* claim plaintiffs pursued. Typically, plaintiffs suing AI

developers for violations of the DMCA have also plead some combination of copyright infringement, trademark infringement or dilution, or unfair competition. Unlike straightforward copyright claims, however, the DMCA does not require the works at issue to be registered, which may have been the reason the *Raw Story Media* plaintiffs pursued only that basis for relief, as they collectively have published over 400,000 news articles and features, and likely have not registered all of them with the U.S. Copyright Office.

What the Future Holds for DMCA Claims against AI Developers. As the *Kadrey and Doe 1* cases demonstrate, plaintiffs may face an uphill battle in holding AI developers liable under the DMCA even before the *Raw Story Media* case demonstrated the challenge of proving standing. While the question of whether generative AI systems are creating derivative works remains hotly contested in litigation around the country, allegations that generative AI systems are generating exact copies of existing works are rarer, and many plaintiffs have also struggled to prove that their work was actually ingested by the model or otherwise lack the required evidence to prove the AI developer knew that the alteration or removal of CMI would result in copyright infringement.

Lacking any one of these allegations was already enough to doom a plaintiff's DMCA claim. *Raw Story Media* now adds standing as another hurdle, as plaintiffs will have to make some nonspeculative allegation that they have been concretely harmed. The easiest way for plaintiffs to do so—proving that the generative AI model in question produced an exact copy of their copyrighted work without the CMI present—is likely to be difficult, as most AI developers have built guardrails in their models to prevent regurgitation of training data.

Plaintiffs have not (yet) been deterred, however: three other suits filed this year by journalists at the Intercept, the Daily News and the Center for Investigative Reporting have included DMCA Section 1202 claims, and many existing suits have active DMCA claims. The next case where we expect a court to consider a Section 1202 claim is *The Intercept Media, Inc. v. OpenAI, Inc. et al*, No. 1:24 CV 01515 (S.D.N.Y. Feb, 28, 2024), which (like *Raw Story Media*) is a suit against an AI developer that only involves allegations of DMCA violations, and the defendant is moving to dismiss for lack of standing and failure to state a claim. While that motion has yet to be decided, the consistent trend in the caselaw so far suggests the motion to dismiss is likely to succeed. Absent significantly stronger allegations by plaintiffs showing that AI models are generating exact copies of copyrighted works from their training data, AI developers are likely to continue defeating DMCA challenges at the early stages of litigation.

* * *

To subscribe to the Data Blog, please [click here](#).

The [Debevoise Data Portal](#) is an online suite of tools that help our clients quickly assess their federal, state, and international breach notification and substantive cybersecurity obligations. Please contact us at dataportal@debevoise.com for more information.

* * *

Please do not hesitate to contact us with any questions.



Megan K. Bannigan
Partner, New York
+1 212 909 6127
mkbannigan@debevoise.com



Christopher S. Ford
Counsel, New York
+1 212 909 6881
csford@debevoise.com



Samuel J. Allaman
Associate, New York
+1 212 909 6026
sjallaman@debevoise.com



Abigail Liles
Associate, New York
+1 212 909 6387
aelliles@debevoise.com



Joshua Plastrik
Associate, New York
+1 212 909 6953
jhplastrik@debevoise.com

This publication is for general information purposes only. It is not intended to provide, nor is it to be used as, a substitute for legal advice. In some jurisdictions it may be considered attorney advertising.