



December 3, 2024

If you have any questions regarding the matters discussed in this memorandum, please contact the following attorneys or call your regular Skadden contact.

Stuart D. Levi

Partner / New York
212.735.2750
stuart.levi@skadden.com

Jack Broitman

Law Clerk / New York
212.735.2084
jack.broitman@skadden.com

This memorandum is provided by Skadden, Arps, Slate, Meagher & Flom LLP and its affiliates for educational and informational purposes only and is not intended and should not be construed as legal advice. This memorandum is considered advertising under applicable state laws.

One Manhattan West
New York, NY 10001
212.735.3000

Digital Millennium Copyright Act Claims in AI-Training Cases – Recent Developments

While the question of fair use has dominated much of the discussion on whether copyrighted material can be used to train AI models, of equal importance are questions involving the application of the Digital Millennium Copyright Act (DMCA) to such activity; specifically whether such training results in the removal of copyright management information, (CMI) in violation of the DMCA. Two recent district court decisions — *Raw Story Media, Inc. v. OpenAI* and *The Intercept Media, Inc. v. OpenAI, Inc.* — shed light on how courts are approaching this issue. In addition, the plaintiffs in *Andersen v. Stability AI*, made a tactical decision to preserve their own DMCA claims. We examine each of these developments below.

Background

Section 1202(b)(1) of the DMCA, which was enacted in 1998, prohibits the intentional removal or alteration of CMI where the removing party knows or has reasonable grounds to know, that it will induce, enable, facilitate or conceal copyright infringement. With respect to text-based materials, CMI is defined to include a work's title, author, copyright owner, and terms and conditions of use that were conveyed with the work. For example, information in a copyright notice or in metadata to a work can constitute CMI. Most courts have interpreted Section 1202(b)(1) as creating a double-scienter knowledge requirement: The defendant (i) must know that CMI has been removed without the authority of the copyright owner or as permitted under law, and (ii) must know or have reasonable grounds to know that such distribution will induce, enable, facilitate or conceal an infringement.

The law was enacted, in part, to address how CMI was often stripped out of works given bandwidth constraints in the late 1990s. Congress noted that preserving CMI would aid in “indicating attribution, creation and ownership” of a work, and would help track and monitor copyright uses to facilitate licensing agreements.

Raw Story Media, Inc. v. OpenAI

Raw Story Media, Inc. and AlterNet Media Inc. (the plaintiffs) are news organizations that collectively own over 400,000 news articles and opinion columns. The plaintiffs brought an action under Section 1202(b)(1) against OpenAI, Inc. and various interrelated organizations that are behind the ChatGPT service, alleging that “thousands” of the plaintiffs’ copyrighted works were used to train ChatGPT, and that in this process any CMI included with those works was removed. The plaintiffs further alleged that there was a substantial likelihood that ChatGPT would generate “verbatim or nearly verbatim” copies of copyrighted works without the applicable CMI, and that the defendant knew the CMI was removed and knew or should have known it would produce responses that would induce, enable, facilitate or conceal infringements in violation of the DMCA. In contrast to most AI training data cases, the *Raw Media* plaintiffs only alleged a DMCA violation and not any copyright infringement claims.

Digital Millennium Copyright Act Claims in AI-Training Cases – Recent Developments

The defendants moved to dismiss, arguing, in part, that the plaintiffs lacked standing because they had failed to properly plead any injury from the alleged removal of the CMI.

On November 7, 2024, U.S. District Judge Colleen McMahon of the U.S. District Court for the Southern District of New York (SDNY) issued an order granting the defendants' motion to dismiss all claims brought for lack of Article III standing. In rendering her decision, Judge McMahon focused on whether the plaintiffs suffered injury-in-fact, a key requirement for Article III standing. For statutory violations, such as the one the plaintiffs alleged, the injury must be concrete and bear a close relationship to a harm traditionally recognized as providing a basis for a claim. The requirement that a plaintiff must establish specific "concrete injury" comes from the Supreme Court's 2021 decision in *TransUnion v. Ramirez*,¹ which held that a statutory right to bring a claim was not sufficient to establish standing in federal court. Plaintiffs must also demonstrate specific "concrete harm" to satisfy Article III's standing requirement.

Here, the plaintiffs argued that their injury bears a "close relationship" to copyright infringement since both DMCA Section 1202(b)(1) and the Copyright Act protect a copyright owner's sole prerogative to decide how future versions of their work may differ from the original. According to the plaintiffs, a DMCA claim is therefore aligned with traditional common law interference with property.

The court disagreed, noting that the DMCA merely protects the integrity of a work's CMI, not how it is used. Specifically, a defendant could copy a work without permission and not be in violation of Section 1202(b)(1) if it kept the CMI intact. Therefore, the plaintiffs' analogy to interference with property was misplaced. The court also cited the legislative history of the DMCA to support the conclusion that the purpose of the DMCA was not to guard against a property-based injury, but rather to insure the integrity of the electronic marketplace for works by preventing fraud and misinformation. The court also noted that the plaintiffs failed to allege any actual adverse effects from the alleged DMCA violation, and, therefore, any harm cannot be considered concrete.

A key omission in the plaintiffs' complaint, according to the court, was any evidence showing that ChatGPT had disseminated (or was likely to disseminate) the plaintiffs' copyrighted work without its CMI.

Judge McMahon also rejected the plaintiffs' argument that it had standing for injunctive relief based on the substantial risk that ChatGPT would generate outputs that were verbatim or near

verbatim copies of the plaintiffs' works. The court noted that under Article III standing requirements, a party has standing to pursue relief when the risk of harm is "sufficiently imminent and substantial." Here, the court agreed with the defendants that given the massive amounts of content on which ChatGPT was trained, coupled with the fact that some of the plaintiffs' works were non-copyrightable facts, meaning that there had not been a showing of a substantial risk of future harm, there was no concrete injury for standing.

The court commented that the real alleged injury for which the plaintiffs were seeking redress was not the exclusion of CMI, but rather the defendants' use of the plaintiffs' articles to train ChatGPT without compensation. The court stated that while there may be a claim for such an injury, it was not to be found within the parameters of Section 1202(b)(1).

While the court granted the plaintiffs leave to amend, Judge McMahon said she was "skeptical" of the plaintiffs' ability to allege a cognizable injury (*i.e.*, that verbatim or near verbatim copies of its works are being generated by ChatGPT).

The Intercept Media, Inc. v. OpenAI, Inc.

Shortly after the decision in *Raw Media*, in *The Intercept Media, Inc. v. OpenAI, Inc.*, Judge Jed Rakoff of the SDNY allowed an AI training case to proceed with respect to its Section 1202(b)(1) claims. The Intercept Media, an investigative news organization, alleged, in part, that its works were used to train ChatGPT and, given certain prompts, the output generated mimicked The Intercept's copyright-protected works without the inclusion of its CMI. The plaintiffs further alleged that the defendants knew or had reason to know that ChatGPT would produce such outputs. While The Intercept had not initially included any actual examples of such outputs in its complaint, it filed an amended complaint that offered three such examples. However, in each of these examples, The Intercept used as input four paragraphs of one of its articles and the first sentence of the fifth paragraph, and prompted ChatGPT to "respond with continuation only." In the three examples provided by The Intercept, ChatGPT completed the next 20 or so words from the articles in question.

The defendants moved to dismiss on numerous grounds, including lack of standing, and subsequently filed a supplemental brief citing the holding in *Raw Media* to support its position. Nonetheless, on November 21, Judge Rakoff issued an order dismissing the defendants' motion to dismiss in part, and allowing the Section 1202(b)(1) claims against OpenAI to survive. Judge Rakoff's full opinion has not yet been issued, so it remains to be seen how the court distinguished its holding from

¹ 141 S.Ct. 2190 (2021).

Digital Millennium Copyright Act Claims in AI-Training Cases – Recent Developments

Judge McMahon’s in *Raw Media*. It is possible that Judge Rakoff found the three examples of partial output text to be sufficiently “concrete” to satisfy the Article III standing requirement and get past the motion to dismiss stage.

Andersen v. Stability AI

In *Andersen v. Stability AI*, an AI training data case involving the unauthorized use of copyrighted images, a California district court recently dismissed the plaintiffs’ DMCA allegations, holding that a claim involving the removal of CMI requires identity between the original work and the copy, which the plaintiffs had failed to establish. However, the issue of whether identity is required for a Section 1202(b)(1) claim is currently on interlocutory appeal to the Ninth Circuit in *Doe 1 v. GitHub, Inc.*² With the appeal pending, and its potential impact on *Andersen*, the plaintiffs in *Andersen* reached an accommodation with the defendants in which the plaintiffs agreed to voluntarily dismiss with prejudice their DMCA claims, in exchange for the defendants agreeing not to challenge future reconsideration of those claims based on their omission from that complaint.

² No. 22-CV-06823-JST, 2024 WL 235217.

Key Takeaways

- DMCA claims have been made in numerous AI training data cases, often in addition to copyright infringement claims. However, if courts adopt Judge McMahon’s approach and find that there is no injury, and therefore no standing, to support these claims, Section 1202(b)(1) claims may be alleged less often unless there is evidence of infringing outputs being generated with the CMI omitted.
- Although *TransUnion* did not involve AI, the Supreme Court’s establishment of a “concrete injury” requirement, even where there is a statutory basis to bring a claim, could impact a number of AI cases based on sections of the Copyright Act, which do not themselves require a showing of actual harm.
- Even where plaintiffs can establish standing, if courts continue to hold that the DMCA claims require identity between the original work and the work that does not include the CMI, it may be difficult for plaintiffs to sustain these claims since AI models typically do not generate identical works.