

Newsrooms vs. neural nets: how courts are handling DMCA claims against genAI

By Sasha S. Rao, Esq., and Todd M. Hopfinger, Esq., Sterne, Kessler, Goldsten & Fox PLLC

AUGUST 27, 2025

The Digital Millennium Copyright Act (DMCA), a 30-year-old tool enacted to address the copyright implications of disrupting technologies, like Napster and Limewire, in the late 1990s, has found new utility in the age of AI. Copyright litigation against GenAI companies using protected works in their training data have often included DMCA claims. But litigants asserting these claims have had mixed results establishing standing and surviving the pleading standard.

Recent cases present an unsettled legal landscape but shed some light on the allegations necessary for viable DMCA claims.

Background

Facing a flood of new internet and digital technology overrunning old copyright paradigms, the DMCA, enacted in 1998, provided updated protections for copyright holders while establishing safe harbor provisions for online and internet service providers to limit their liability for copyright infringement.

The Act includes measures to prevent circumvention of technical solutions to restrict access to copyrighted materials and to stop trafficking in devices whose purpose is to circumvent such measures. Among others, the DMCA includes a provision, Section 1202(b)(1), to penalize the unauthorized, intentional removal or alteration of copyright management information (CMI), i.e., information conveyed with copies of a work, including its title, authorship, copyright ownership, and terms and conditions for use.

DMCA claims differ from traditional copyright claims. Rather than show the unauthorized reproduction of copyrighted materials, DMCA plaintiffs must show the intentional removal of CMI from copyrighted works and/or distribution of those works without CMI. The DMCA is thus relevant in the gen-AI economy where such information can get stripped during the processing of training data and is unlikely to accompany any outputs. A successful plaintiff who has shown a DMCA violation could obtain actual damages and any additional profits of the violator or statutory damages.

Another DMCA section, i.e., 1202(b)(3), prohibits the distribution of copyrighted work that would induce, enable, facilitate, or

conceal infringement while knowing CMI has been removed. But these have been far less successful in surviving the initial stages of litigation. Thus, this article does not address that provision.

DMCA standing obstacles and opportunities

Fundamental to any action is plaintiffs' ability to establish standing. Under the Supreme Court's *TransUnion LLC v. Ramirez* decision, the requisite "concrete harm" for Article III standing must bear "a close relationship to a harm traditionally recognized as providing a basis for a lawsuit in American courts." But the problem facing courts is that there is very little that is "traditional" about AI.

Recent cases present an unsettled legal landscape but shed some light on the allegations necessary for viable Digital Millennium Copyright Act claims.

How courts characterize the "harm" in genAI cases has dictated DMCA claims' survival. Two Southern District of New York cases against OpenAI are illustrative: *Raw Story Media, Inc. et al. v. OpenAI, Inc. et al.* (decided by Judge Colleen McMahon) and *The Intercept Media, Inc. v. OpenAI, Inc.* (decided by Judge S. Rakoff).

Both *Raw Story* and *Intercept Media* involve OpenAI's use of news organizations' materials to train its large-language model (LLM). Both sets of plaintiffs alleged that OpenAI had intentionally omitted CMI from the articles included in its training data. Yet, these cases had opposite outcomes.

In *Raw Story*, Judge McMahon found no injury-in-fact had been plausibly alleged. In assessing the alleged injury, the court observed that the harm alleged was just the unauthorized removal of CMI from the copyrighted works, not the dissemination of those works without the CMI.

The plaintiffs argued that the removal of CMI bears “close relationships” to copyright infringement and interference with property. The court disagreed. Judge McMahon concluded that the harm claimed, which she characterized as the “mere removal of identifying information from a copyrighted work-absent dissemination,” was not a traditional harm.

Moreover, because the plaintiffs had shown no instance (nor a “substantial risk”) of ChatGPT actually outputting copyrighted text, any harm was too abstract and speculative. In fact, given the “massive” size of ChatGPT’s training corpus, the judge found “the likelihood that ChatGPT would output plagiarized content from one of plaintiffs’ articles seemed remote.”

DMCA claims differ from traditional copyright claims. Rather than show the unauthorized reproduction of copyrighted materials, DMCA plaintiffs must show the intentional removal of CMI from copyrighted works and/or distribution of those works without CMI.

By contrast, in *Intercept Media*, Judge Rakoff found that the plaintiff had alleged a concrete injury and thus had standing. Even though the specific right to CMI is a modern, DMCA-created right, the court reasoned that removal of attribution implicates “the same kind of property-based harms traditionally actionable in copyright.” This historical harm is the injury to the author’s IP rights and the incentives to create. The loss of control over how plaintiff’s work is presented (authorship info stripped) could lead to uncredited copying, which the court found analogous to classic copyright harms.

A 2025 decision by yet a third judge out of the Southern District of New York has followed the more expansive view of injury of *Intercept Media*. In *The New York Times v. Microsoft Corp. et al.*, the court again considered news organizations’ DMCA claims against genAI companies’ (OpenAI and Microsoft) using their materials for LLM training. Citing *Intercept Media*, the Court found standing, reasoning that “[f]or both DMCA and traditional copyright infringement claims, the harm involves an injury to ‘an author’s property right in his original work of authorship.’”

In addition to *Intercept Media* and *Raw Story* courts’ differing views of injury for DMCA claims, these decisions also reveal a split in the relevance of an LLM’s output. The *Raw Story* court suggests that allegations of dissemination and “any actual adverse effects stemming from [an] alleged DMCA violation” could have changed the result. But in *Intercept Media*, the

court held that the injury “does not require publication to a third party,” rejecting OpenAI’s argument that no one besides the plaintiff itself had seen any offending output.

DMCA pleading obstacles and opportunities

Along with standing, DMCA claims have unique challenges for the sufficiency of their pleadings. While most DMCA violations concern LLM inputs, the nature of their outputs carry significant weight at the pleadings stage. Specifically, recent cases have established an “identity” requirement that LLM outputs must be identical to the accused training data for DMCA liability from CMI-removal during LLM training.

These decisions have also required a “double scienter” showing: the defendant must (1) knowingly remove or alter CMI, and (2) know or have reason to know that the act will facilitate infringement.

A 2024 decision in *Andersen et al. v. Stability AI Ltd. et al* is illustrative. There, plaintiff artists alleged that Stability AI and other AI companies used the artists’ works as training data for their image-generation model Stable Diffusion and removed the works’ CMI while ingesting them into the data.

The plaintiffs supported their allegations by comparing some of plaintiffs’ images used as training images and that contained CMI with images created through the AI-image prompts. And plaintiffs specifically identified the allegedly stripped CMI. Plaintiffs also argued that defendant engaged in knowing CMI removal because of how the AI model worked, how the training images are used, and plausible allegations regarding Stability AI’s funding of the training dataset.

Despite this comprehensive showing, the court dismissed all Section 1202(b) claims, holding that plaintiffs failed to allege that any AI-generated outputs were *identical* to the plaintiffs’ works. The court adopted a narrow reading of Section 1202(b): CMI must be removed from an identical copy of the work to trigger liability.

Yet, Section 1202(b)(1) claims remain viable where plaintiffs allege concrete acts of CMI removal during training and provide contextual facts suggesting the AI developer knew this could facilitate infringement. In *The New York Times v. Microsoft Corp. et al.*, the court allowed certain Section 1202(b) (1) claims to survive the pleading stage for certain defendants based on the high level of detail in their complaints. In particular, plaintiff CIR explained in detail how defendants removed CMI from the training data and that defendants knew their tools would remove that CMI.

By contrast, the court dismissed The New York Times’ Section 1202(b) claim, finding that this defendant’s complaint failed to include “any specific detail on how CMI was allegedly removed during the training process” and that its allegation that defendants’ removal of CMI in its LLM-training process is “by design” was too conclusory. A notable distinction from *Stability AI* — albeit one not relied upon by this court — is that the LLM at issue here could regurgitate plaintiffs’ work,

which could more easily meet the identity requirement and demonstrate concrete injury-in-fact. By contrast, the output in *Stability AI* was admittedly not identical to the training images but rather generated in the same style.

Conclusion and takeaways

The DMCA affords plaintiffs another cause of action against the unauthorized use of their protected works by GenAI

companies. The discussed cases present an evolving legal landscape with certain key principles emerging. First, standing thresholds may vary by court and that court's characterization of the harm of a DMCA violation. Second, Section 1202(b)(1) claims are an uphill climb with the identity and double-scienter requirements but are nonetheless viable with highly detailed allegations and attention to outputs.

About the authors



Sasha S. Rao (L) is counsel in **Sterne, Kessler, Goldstein & Fox's** trial and appellate practice group, in Washington, D.C. He has been involved in various matters before federal district courts, the International Trade Commission, and the U.S. Court of Appeals for the Federal Circuit. His experience spans from initial fact investigation through trial in a wide range of technologies, including artificial intelligence (AI), pharmaceuticals, materials science, chemical arts, and automotive technology. He can be reached at SRao@sternekessler.com. **Todd M. Hopfinger** (R) is a director in the firm's electronics practice group in Washington, D.C. He specializes in strategic

intellectual property (IP) counseling and patent prosecution for technology companies at all development stages, from innovative startups to major global enterprises. His areas of expertise include cutting-edge fields such as artificial intelligence (AI), bioinformatics, data science, blockchain, 5G, consumer technology, transportation and logistics (including autonomous vehicles), digital advertising, and cybersecurity. He can be reached at thopfinger@sternekessler.com.

This article was first published on Reuters Legal News and Westlaw Today on August 27, 2025.