Beyond Infringement: Rethinking DMCA § 1202 for Generative AI
78 STAN. L. REV. (forthcoming 2026)

NOTE

Beyond Infringement: Rethinking DMCA § 1202 for Generative AI

LARISSA BERSH[†]

Abstract. With the emergence of large language models (LLMs) like ChatGPT, scholars and courts have fervently debated whether LLMs' training on and reproduction of copyrighted materials amounts to fair use. But in a recent series of cases, a lesser-known challenge to LLMs has reared its head: § 1202 of the Digital Millenium Copyright Act. This provision requires that when a work is copied, its associated copyright management information (CMI)—such as its license or terms of use agreement—is copied with it. The requirement was originally intended to modernize copyright for the Internet by ensuring all users would be aware of the terms of their use. Now, § 1202's unintended overbreadth threatens to block LLM development and use as it swallows questions of infringement and fair use entirely.

This Note argues that § 1202 is broader than traditional copyright infringement in three critical respects: It allows liability without any showing of copyrightability, provides no fair use defense, and permits disproportionate statutory damages. While § 1202 includes an intent requirement, that requirement is so minimal that it fails to meaningfully limit the statute's reach—especially in the LLM context, where the act of violating § 1202 may itself satisfy the mental requirement.

In response to these concerns, this Note proposes that courts adopt an identicality requirement for § 1202 claims against LLMs. This requirement would cabin liability to outputs that exactly match training data—cases where the removal or alteration of CMI is both clear and technically avoidable. The approach mirrors Congress's existing accommodation of technological limitations for broadcasters in cases of technical infeasibility or financial hardship. A similar understanding for LLMs would preserve § 1202's core purpose, resolve a growing district court split, and ensure the most consequential copyright question raised by generative artificial intelligence is answered on its merits—not sidestepped by a statute never designed to decide it.

[†] [Acknowledgment Text]

Table of Contents

Introduction 3

I. The DMCA's Purpose: Maintaining Copyright Liability in a Digital World 9

II. The DMCA in Practice: Liability Explodes 12
A. Weakened Standard for Proving Copying 13
B. No Fair Use Defense 24
C. Heightened Statutory Damages 27
D. The Intent Requirement Fails to Cabin 30

III. Applying the DMCA to LLMs 34

Conclusion 53

for proving copying under infringement, the copyrightability of the underlying work is not considered.[113]

The courts implementing an identicality requirement consist mostly of district courts within the Ninth Circuit.[114] These courts have held that an allegedly infringing work must be an identical copy of the original to violate § 1202.[115] But even among individual courts, whether the identicality requirement is applied has been inconsistent. For example, three judges in the Northern District of California came to three very different conclusions: one found that an LLM distributing outputs without CMI was insufficient to support a DMCA claim where those outputs were derivative works rather than identical copies of inputs;[116] one found in contrast that a § 1202(b) claim *survived* a

[113] A smattering of courts has considered copyrightability in an unsystematic and ungeneralizable fashion. One court focused on the copyrightability inquiry within the context of class certification, finding that the plaintiffs did not sufficiently establish the common ownership necessary to certify their class because they did not suggest a manner for proving their copyright ownership. *See Schneider v. YouTube, LLC*, 674 F. Supp. 3d 704, 723-724 (N.D. Cal. 2023). Another court positioned the inquiry within the scienter requirement, finding that the plaintiffs failed to show the necessary scienter because the works they alleged were the subject of the § 1202 violation were owned by someone else entirely. *See Munro v. Fairchild Tropical Botanic Garden, Inc.*, WL 452257, at *11 (S.D. Fla. Jan 13, 2022). Another court similarly focused on ownership rather than the extent of the work's ability to be copyrighted. Finding "issues of material fact as to which copyrights [plaintiff] owns and may enforce." *Sanborn Libr. LLC v. ERIS Info, Inc.*, WL 1744630, at *20 (S.D.N.Y. Mar. 25, 2024). Yet another court imported a copyrightability requirement seemingly by mistake, stating that "[t]o establish copyright infringement for all three counts," including a copyright infringement count and two § 1202 counts, "plaintiff must show '(1) ownership of a valid copyright, and (2) copying of constituent elements of the work that are original.'" *Say It Visually, Inc. v. Real Est. Educ. Co., Inc.*, WL 933951, at *8 (N.D. Ill. Mar. 27, 2025). In support of this contention, the court cited a case that is entirely about copyright infringement and does not mention § 1202 once. *Id.* The closest a court has come to a proper acknowledgement of the need for a copyrightability analysis was *Raw Story Media*'s statement that "[p]laintiffs have nowhere alleged that the *information* in their articles is copyrighted." *Raw Story Media, Inc. v. OpenAI, Inc.*, 756 F. Supp. 3d 1, 7 (S.D.N.Y. 2024). However, the court is referring to LLMs' "[synthesis of] the relevant information" from its training data—it seems to suggest that LLMs inherently copy unprotectable aspects of works. *Id.* This is distinct from an inquiry into the copyrightability of the training data itself.

[114] *Id.* ("[T]he Court finds that . . . Section 1202(b) claims require that copies be 'identical.'"); *Advanta-STAR Auto. Rsch. Corp. of Am. v. Search Optics, LLC*, 672 F. Supp. 3d 1035, 1057 (S.D. Cal. 2023) ("Plaintiff has not plausibly alleged that Defendants distributed identical copies of Plaintiff's comparison. Accordingly, Plaintiff fails to state a claim for a violation of Section 1202 of the DMCA."); *O'Neal v. Sideshow, Inc.*, 583 F. Supp. 3d 1282, 1287 (C.D. Cal. 2022) ("[Defendants] argue that [Plaintiff's] claim fails as a matter of law because removal of CMI requires the work used by a defendant to be identical … The Court agrees."); *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-1931-DMG (EX), 2020 WL 5991503, at *6 (C.D. Cal. Aug. 14, 2020) ("Defendant did not make *identical* copies of Plaintiff's works and then remove the engraved CMI. In such cases, even where the underlying works are similar, courts have found that no DMCA violation exists where the works are not identical.").

[115] *Id.*

[116] Tremblay v. OpenAI, Inc., 716 F. Supp. 3d 772, 780 (N.D. Cal. 2024) ("Under the plain language of the statute, liability requires distributing the original 'works' or 'copies of [the]

motion to dismiss where the allegedly infringing work was "a derivative of [plaintiff's] copyrighted source code" (and therefore presumably not identical);[117] and one found that the "split in authority" on this issue contributed to an inability to dismiss on the basis of identicality.[118]

Even those courts that have seemingly aligned around the existence of an identicality standard have failed to coalesce around a specific proper application. Some cases have suggested the relevance of an allegedly infringing work's status as a derivative work in determining whether it is identical to the original,[119] with one court explaining that derivative works cannot violate the DMCA because they are not an "'identical copy' of the original."[120] Another court required perfect identicality, dismissing a § 1202(b) claim for lack of identicality where plaintiffs alleged a "functionally equivalent," "nearly verbatim reproduction" with only "a few 'cosmetic' differences in word choice."[121] Other district courts have agreed, finding that plaintiffs failed to state a § 1202(b) claim where works of visual art "were modified or altered"—even where those works "were not altered in substance, but rather in opacity and position."[122] Different still, another court determined lack of identicality by visually analyzing side-by-side images, providing no further details about the analysis.[123] Yet another court stated more generally that § 1202 only applies "to the removal of copyright management information on a plaintiff's product or original work."[124] It is unclear whether, for example, a work with alterations to its form but not its substance would qualify as the plaintiff's original work under this standard.

works.' 17 U.S.C. § 1202(b)(3). Plaintiffs have not alleged that Defendants distributed their books or copies of their books. Instead, they have alleged that "every output from the OpenAI Language Models is an infringing derivative work" without providing any indication as to what such outputs entail – i.e., whether they are the copyrighted books or copies of the books. That is insufficient to support this cause of action under the DMCA.").

[117] Splunk Inc. v. Cribl, Inc., 662 F. Supp. 3d 1029, 1054 (N.D. Cal. 2023) ("The Court may later determine . . . that the [code] was not, in fact, a derivative of [plaintiff's] copyrighted source code. For now, however, the CMI is sufficiently pleaded to survive a motion to dismiss.").

[118] Beijing Meishe Network Tech. Co. v. TikTok Inc., No. 23-CV-06012-SI, 2024 WL 3522196, at *9 (N.D. Cal. July 23, 2024).

[119] Frost-Tsuji Architects v. Highway Inn, Inc., 2015 WL 263556, at *4 (D. Haw. Jan. 21, 2015) (finding that a derivative use of plaintiff's work, rather than a use of the original work, cannot uphold a CMI removal claim under § 1202(b)).

[120] Real World Media LLC v. Daily Caller, Inc., 744 F. Supp. 3d 24, 40 (D.D.C. 2024).

[121] Doe 1 v. GitHub, Inc., No. 22-CV-06823-JST, 2024 WL 235217, at *9 (N.D. Cal. Jan. 22, 2024).

[122] *O'Neal v. Sideshow, Inc.*, 583 F. Supp. 3d 1282, 1287 (C.D. Cal. 2022).

[123] *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-1931-DMG (EX), 2020 WL 5991503, at *6 (C.D. Cal. Aug. 14, 2020).

[124] *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999).