



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 Millennium 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—be copied with it. This requirement was originally intended to modernize copyright for the internet by ensuring that 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 posits that § 1202 is broader than traditional copyright infringement doctrine in three critical respects: It imposes liability without any showing of copyrightability, provides no fair use defense, and permits disproportionate statutory damages. Although § 1202 includes an intent requirement, courts have applied it so minimally that it fails to constrain the statute's reach—especially in the LLM context, where the mere act of violating § 1202 may itself suffice to establish intent.

To restore meaningful limits to § 1202, this Note proposes that courts adopt an identity requirement for § 1202 claims against LLMs. The 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. This approach mirrors Congress's existing accommodation of industry 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 that the most consequential copyright question raised by generative artificial intelligence is answered on its merits—not sidelined by a statute never meant to decide it.

* Stanford Law School, J.D. 2025. Thank you to Phil Malone, Nina Srejovic, Braden Crimmins, and the Juelsgaard Intellectual Property and Innovation Clinic for the teamwork, inspiration, and ideation that led to this Note. Thank you also to the outstanding editors of the *Stanford Law Review*. All errors are my own.

Table of Contents

Introduction669

I. The DMCA’s Purpose: Maintaining Copyright Liability in a Digital World675

II. The DMCA in Practice: Liability Explodes678

 A. Weakened Standard for Proving Copying680

 B. No Fair Use Defense692

 C. Heightened Statutory Damages696

 D. The Intent Requirement Fails to Cabin Liability698

III. Applying the DMCA to LLMs703

IV. Bridging Purpose and Practice716

Conclusion723

Introduction

Commentators estimate that large language models (LLMs) like ChatGPT and Claude could add trillions of dollars in value to the global economy.¹ Meanwhile, the appetite of these systems is growing exponentially as they require access to more and more copyrighted works as training data.²

Unsurprisingly, this tension between economic and authorship interests has spawned a wealth of legal scholarship and public commentary.³ On the legal side, the debate centers around whether training LLMs on copyrighted material is permissible as a “fair use.”⁴ Some scholars vehemently argue that LLM training is a transformative fair use and is therefore permitted even if it would otherwise constitute infringement.⁵ Others are just as emphatic in arguing the opposite.⁶ Courts have begun to weigh in,⁷ with a growing number of lawsuits poised to shed light on the issue.⁸ Non-academic commentators—

-
1. See, e.g., MICHAEL CHUI ET AL., MCKINSEY & CO., *THE ECONOMIC POTENTIAL OF GENERATIVE AI: THE NEXT PRODUCTIVITY FRONTIER* 3 (2023), <https://perma.cc/3X37-CN5V>.
 2. See, e.g., Cade Metz, Cecilia Kang, Sheera Frenkel, Stuart A. Thompson & Nico Grant, *How Tech Giants Cut Corners to Harvest Data for A.I.*, N.Y. TIMES (updated Apr. 8, 2024), <https://perma.cc/GFD2-G9ZK>; Tom B. Brown et al., *Language Models Are Few-Shot Learners*, 3 PROCS. 34TH INT’L CONF. ON NEURAL INFO. PROCESSING SYS., art. 159, at 14 (2020), <https://perma.cc/MX4C-VQSG>.
 3. See, e.g., William D. Cohan, *AI Is Learning from Stolen Intellectual Property. It Needs to Stop*, WASH. POST (Oct. 19, 2023), <https://perma.cc/7GMA-A6PQ>; Robert Harington, *A Dissonance of Ideals: Openness, Copyright, and AI*, SCHOLARLY KITCHEN (Nov. 25, 2024), <https://perma.cc/A8J5-FMRP>; Adam Buick, *Copyright and AI Training Data—Transparency to the Rescue?*, 20 J. INTELL. PROP. L. & PRAC. 182, 182 (2025).
 4. See Keith Kupferschmid, *Insights from Court Orders in AI Copyright Infringement Cases*, COPYRIGHT ALL. (Dec. 12, 2024), <https://perma.cc/35BD-YPLY> (“[I]t should surprise no one that the primary issue in virtually all of the pending AI-copyright cases being litigated in federal courts today is whether the ingestion of an unlicensed copyrighted work for training of a generative artificial intelligence (GAI) model qualifies as fair use.”).
 5. See, e.g., Elizabeth Spica, *Public Interest, the True Soul: Copyright’s Fair Use Doctrine and the Use of Copyrighted Works to Train Generative AI Tools*, 33 TEX. INTELL. PROP. L.J. 67, 70, 72 (2024).
 6. See, e.g., Jacob Alhadeff, Cooper Cuene & Max Del Real, *Limits of Algorithmic Fair Use*, 19 WASH. J.L. TECH. & ARTS 1, 47 (2024) (finding fair use unlikely for text-to-image generative AI where the original artist experiences market harm).
 7. Recently, the District Court for the District of Delaware found in favor of plaintiffs, rejecting the fair use defense on summary judgment. Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc., 765 F. Supp. 3d 382, 401 (D. Del. 2025); see also *An Early Win for Copyright Owners in AI Cases as Court Rejects Fair Use Defense*, DEBEVOISE & PLIMPTON (Feb. 14, 2025), <https://perma.cc/4YF3-9UV8> (reporting on the Thomson Reuters case).
 8. *Status of All 39 Copyright Lawsuits v. AI (Feb. 18, 2025): Judge Bibas Rejects Fair Use in AI Training in Stunning Reversal*, CHAT GPT IS EATING THE WORLD (Feb. 19, 2025), <https://perma.cc/AH4T-JKPY>; Theresa M. Weisenberger, Diana C. Milton & Harrison
footnote continued on next page

from tech executives to public interest advocates—have weighed in as well.⁹ Indeed, the issue has incited political tension, with some suggesting that the firing of former Copyright Office head Shira Perlmutter was linked to the Office’s publication of a contentious report questioning whether fair use applies to LLMs’ training on copyrighted data.¹⁰

But while the spotlight has been on fair use and copyright infringement, a lesser-known legal tension has emerged, coming to light in *Doe v. GitHub, Inc.*¹¹ Unlike most AI-copyright lawsuits, *GitHub* does not allege infringement at all.¹² Instead, the plaintiffs invoke § 1202 of the Digital Millennium Copyright Act (DMCA).¹³

The DMCA is a federal copyright statute enacted in the early days of the internet to prevent digital piracy.¹⁴ Prior to the DMCA’s enactment in 1998, the Copyright Act of 1976 was unable to effectively address the intellectual property challenges associated with rapid advances in technology.¹⁵ The

A. Enright, *Case Tracker: Artificial Intelligence, Copyrights and Class Actions*, BAKERHOSTETLER, <https://perma.cc/ZST4-MMSG> (archived Dec. 30, 2025).

9. See Angela Yang, *OpenAI Urges U.S. to Allow AI Models to Train on Copyrighted Material*, NBC NEWS (Mar. 13, 2025, 1:55 PM PDT), <https://perma.cc/5Z5E-KD4B> (tech executives); Tori Noble, *AI and Copyright: Expanding Copyright Hurts Everyone—Here’s What to Do Instead*, ELEC. FRONTIER FOUND. (Feb. 19, 2025), <https://perma.cc/44VL-FMHC> (public interest advocates).
10. Wes Davis, *Trump Fires Head of Copyright Office Two Days Following Report That AI Training May Not Be Fair Use*, VERGE (May 11, 2025, 9:15 AM PDT), <https://perma.cc/3B27-4D9C>; see Tori Noble, Mitch Stoltz & Corynne McSherry, *The U.S. Copyright Office’s Draft Report on AI Training Errs on Fair Use*, ELEC. FRONTIER FOUND. (May 15, 2025), <https://perma.cc/7EXD-8UZ9>.
11. No. 22-cv-06823, 2024 WL 235217 (N.D. Cal. Jan. 22, 2024), *argued*, No. 24-7700 (9th Cir. Feb. 11, 2026).
12. See Second Amended Complaint ¶¶ 234-35, *GitHub*, 2024 WL 235217 (No. 22-cv-06823), ECF No. 200 (alleging a violation of the Digital Millennium Copyright Act (DMCA) and breach of contract).
13. Digital Millennium Copyright Act, 17 U.S.C. § 1202; Second Amended Complaint, *supra* note 12, ¶ 205.
14. Zoe Carpou, *Robots, Pirates, and the Rise of the Automated Takedown Regime: Using the DMCA to Fight Piracy and Protect End-Users*, 39 COLUM. J.L. & ARTS 551, 556 (2016) (“According to the statute’s legislative history, Congress had two goals in creating the DMCA These goals arose in part out of Congress’s concern about widespread piracy made possible by the rise of digital media.”).
15. Michelle A. Ravn, Note, *Navigating Terra Incognita: Why the Digital Millennium Copyright Act Was Needed to Chart the Course of Online Service Provider Liability for Copyright Infringement*, 60 OHIO ST. L.J. 755, 758-60 (1999) (explaining how the internet’s enablement of the easy copying and transmission of copyrighted content led to questions about the liability of online service providers); Haley Bridget McCullough, Note, *Closing the Gap Between Copyright Management Information and Metadata: A Critical Analysis of Section 1202 of the Digital Millennium Copyright Act and a Proposal for Sound Recording Standards*, 24 U. DENV. SPORTS & ENT. L.J. 75, 88 (2021) (“It is well settled that
footnote continued on next page

internet allowed for millions of exact copies of copyrighted works to be shared across the globe in seconds at minimal cost and with “virtually no degradation in quality.”¹⁶ As creative work spread across the internet through social networks, copies of copies propagated in untraceable patterns. These works may once have contained author-imposed conditions of duplication known as copyright management information (CMI), such as requiring derivative works to credit the original author or remain accessible to the public. But once a work was copied without its CMI, subsequent viewers had no way of knowing what uses were permissible. The author’s intended restrictions were obscured, making inadvertent infringement inevitable and effective enforcement impossible. In response, the Clinton Administration proposed a series of legal protections above and beyond traditional copyright—a proposal that eventually culminated in the DMCA.¹⁷

Under DMCA § 1202, if a license contains CMI, that CMI must be preserved when a work is copied to enable future users to comply with the work’s conditions for use.¹⁸ CMI can take the form of a terms-of-service agreement, an open-source license, a work’s title, the name of the author or copyright owner, or other identifying information.¹⁹ As such, § 1202 protections do not protect the works themselves, but rather the digital management system that protects copyrights.²⁰ Thus, DMCA and traditional copyright infringement claims will often be brought together, each also applies where the other does not. A work can infringe copyright without violating § 1202, and a work may violate § 1202 without infringing copyright.²¹ For example, unlawfully copying a work may provide a basis for an infringement claim, but if the defendant keeps the work’s CMI intact, the plaintiff will not

before the DMCA was enacted in 1998, the Copyright Act of 1976 never adequately addressed advances in technology nor the exponential growth of the Internet.”).

16. McCullough, *supra* note 15, at 90.

17. S. REP. NO. 105-190, at 65-66 (1998) (describing the Clinton Administration’s release of a report recommending copyright protections that were later “echoed by the Digital Millennium Copyright Act”).

18. Maria Crusey, Note, *Section 1202(b) and AI: Implications for Copyright Infringement Lawsuits and Considerations for Digital Creators*, 72 J. COPYRIGHT SOC’Y 480, 485 (2025) (“Both Section 1202(a) and Section 1202(b) are rooted in the premise that removal of CMI can further copyright infringement by enabling greater access to works that are not clearly protected under copyright or associated with an author.”).

19. 17 U.S.C. § 1202(c).

20. TOM W. BELL, *INTELLECTUAL PRIVILEGE: COPYRIGHT, COMMON LAW, AND THE COMMON GOOD* 30 (2014).

21. McCullough, *supra* note 15, at 94 (“What may seem like an obvious case of copyright infringement may not necessarily be a violation of Section 1202.” (capitalization altered)).

have a viable § 1202 claim.²² Conversely, if a copy is a fair or licensed use but removes or alters CMI, the plaintiff may have a basis for a § 1202 claim but not for copyright infringement.²³

While some of § 1202's breadth may have been intentional,²⁴ much of its expansiveness appears accidental. Unanswered questions and contradictory interpretations have left courts without clear guardrails.²⁵ For example, courts and commentators disagree on whether fair use is an applicable defense to a § 1202(b) claim and on what the proper standard is to qualify as a "copy" of a work under § 1202(c), among other disputes.²⁶

Though § 1202 was created to accommodate changes brought about by the digital age, its unpredictability and breadth are anything but accommodating for the new technologies of today. Indeed, for generative artificial intelligence (AI), § 1202 threatens to hamper creativity, not foster it.

Models like ChatGPT are trained on hundreds of thousands of copyrighted works.²⁷ Reading § 1202 on its face, the LLM training process is plainly

-
22. See 4 MELVILLE B. NIMMER & DAVID NIMMER, *NIMMER ON COPYRIGHT* § 12A.10 (LexisNexis 2025) ("Admittedly, an unauthorized film may infringe the copyright in the novel on which it is based, but crediting the film to someone other than the novelist does not seem to rise to the level of false copyright management information." (footnotes omitted)).
 23. See Peter Henderson, Xuechen Li, Dan Jurafsky, Tatsunori Hashimoto, Mark A. Lemley & Percy Liang, *Foundation Models and Fair Use*, 24 J. MACH. LEARNING RSCH. 21 (2023) ("Even if model creators rely on data under open-source licenses, there may be other issues that arise from removing licensing information. DMCA § 1202 creates liability if someone intentionally removes [CMI] or knowingly distributes content with the CMI removed.").
 24. Megan Keenan, *Copyright Management Information (CMI) as a Tool to Protect Indigenous Cultural Works*, 12 N.Y.U. J. INTELL. PROP. & ENT. L. 413, 421 (2023) (referring to the statute's "broad definition of CMI").
 25. 4 NIMMER & NIMMER, *supra* note 22 § 12A.10 (discussing the confusion created by § 1202's unanswered questions).
 26. See *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417, 2025 WL 1786418, at *1 (N.D. Cal. June 27, 2025) (holding that fair use precludes the finding of a § 1202 violation); *Murphy v. Millennium Radio Grp. LLC*, No. CIV.A. 08-1743, 2015 WL 419884, at *5 (D.N.J. Jan. 30, 2015) (finding that fair use does not preclude § 1202 liability because it "does not mean that Defendants could not have intended to abet an infringement" under § 1202); see also Karen A. Chesley, *Calculating Damages Under the Digital Millennium Copyright Act: How Far Should Courts Go when Multiplying Statutory Awards?*, 57 J. COPYRIGHT SOC'Y U.S.A. 17 (2010) (describing one court's anomalous finding that "damages should be aggregated based on each view" of a violating video, "ignor[ing] the fact that such an analysis contravenes the requirement of a deliberate, voluntary 'act'").
 27. See Daniel Rodríguez Maffioli, *Copyright in Generative AI Training: Balancing Fair Use Through Standardization and Transparency* 4 (Oct. 18, 2023) (unpublished manuscript), <https://perma.cc/4DVG-SS9X> (referring to the "copious amounts of copyrighted data gathered for training [AI] models"); Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 745 (2021) ("Creating a training set of millions of examples almost always requires, first, copying many more millions of images, videos, *footnote continued on next page*

incompatible with the statute's requirements. Section 1202(b) maintains that "no person shall . . . intentionally remove or alter any copyright management information,"²⁸ but separating CMI from its source is a necessary part of the LLM training process.²⁹ Though LLMs could attempt to show users the CMI for the works that contributed to the output, presenting the CMI for hundreds of thousands of works is unhelpful, and attributing outputs to specific sections of training data remains a technical challenge.³⁰ Whether this training is a fair

audio, or text-based works. Those works are almost all copyrighted." (footnote omitted)).

28. 17 U.S.C. § 1202(b).

29. Bruce E. Boyden, *Generative AI and IP Under US Law*, in *THE CAMBRIDGE HANDBOOK OF GENERATIVE AI AND THE LAW* 270, 278-79 (Mimi Zou, Cristina Poncibò, Martin Ebers & Ryan Calo eds., 2025) ("Obviously the training process for a generative AI model is not intended to preserve a work in its entirety, including any CMI contained in the work itself or a file's metadata. . . . But training an AI model by its nature scrambles the plaintiff's works into a disparate set of associated data points."); Pamela Samuelson, Christopher John Sprigman & Matthew Sag, Comment Letter on Notice of Inquiry on Artificial Intelligence and Copyright 8 (Oct. 30, 2023), <https://perma.cc/227E-T575> ("The generative AI training process extracts information from millions or billions of works and, in the process, disassembles or tokenizes their elements to construct a very different representation in the models."); see also Crusey, *supra* note 18, at 483 ("Removal of CMI and other forms of identifying information is necessary for the effective creation and development of generative AI systems."). Some courts have suggested that because LLMs sometimes output copies of works with the appropriate CMI, those LLMs do not systematically remove CMI. See *Concord Music Grp. v. Anthropic PBC*, No. 24-cv-03811, 2025 WL 1487988, at *7 (N.D. Cal. Mar. 26, 2025) ("The Court notes that Publishers have alleged that some outputs do include attribution for lyrics. These allegations arguably undermine Publishers' theory that Anthropic intentionally removes CMI by design because they suggest that CMI was not removed across the board—intentionally or otherwise—during Claude's training process." (emphasis omitted) (citations omitted)). This argument is insufficient for two reasons. First, it only addresses potential CMI removal during output, not training. LLM training involves the scrambling, disassembling, and extraction of all parts of the training data. See Boyden, *supra*, at 14; Samuelson et al., *supra*, at 8. If LLMs can violate § 1202 by removing CMI during training, then the CMI referenced in *Concord* was systematically removed in this scrambling and only later put back together. Second, outputting even correct CMI constitutes effective removal because a user has no way of knowing whether the CMI is correct without guessing that the output is a copy and locating the original—putting the user in the same position they would be in absent the CMI output and making § 1202 powerless to aid traditional copyright.

30. Zeynep Ülkü Kahveci, *Attribution Problem of Generative AI: A View from US Copyright Law*, 18 J. INTELL. PROP. L. & PRAC. 796, 798-99 (2023) ("Giving attribution to works that are used to create a certain output requires additional extensive work by the AI systems as it is not easy to determine which specific works are used for a specific output."); see also Henderson et al., *supra* note 23, at 19 ("[F]oundation models rarely, if ever, provide proper attribution. . . . Indeed, in many cases, it can be difficult to determine which training examples actually contributed to a given generation."). Though LLMs have outputted CMI, so long as the developer is operating in good faith, the function of an LLM is to learn from its training data, not to copy it. As such, an instance of outputted CMI typically represents not attached, associated information, but the LLM's guess of

footnote continued on next page

use under traditional copyright infringement has been addressed on narrow facts at the district-court level, but further attempts to answer the question will be moot if the training process remains independently punishable under § 1202.³¹ Courts generally have not considered fair use to be an applicable defense under that title, and § 1202 statutory damages can be exorbitant.³²

Courts are split on whether § 1202 requires that an output be identical to the copyrighted source or merely similar. This disagreement—most notably among district courts within the Ninth Circuit, but even within the Northern District of California itself³³—played out in *Doe v. GitHub*, a case in which the Northern District adopted an identity requirement.³⁴ Citing the split, the Ninth Circuit granted an interlocutory appeal to resolve the issue.³⁵

Though narrow in framing, the issue has sweeping implications for LLM providers who may face § 1202 liability for their models' outputs. If identity is required, liability would be limited to outputs that clearly reproduce training data—cases where malintent is clear, CMI removal was technologically avoidable, and future infringement is likely. But if similarity is enough, § 1202 becomes a blunt instrument: It could impose billions in statutory damages for even the most well-intentioned, unavoidable, and fair uses.

what text tends to appear in the given context. See, e.g., Albert Ziegler, *GitHub Copilot Research Recitation*, GITHUB BLOG (updated Aug. 16, 2022), <https://perma.cc/7NG2-DVF7> (“In one instance, GitHub Copilot suggested starting an empty file with something it had even seen more than a whopping 700,000 different times during training—that was the GNU General Public License.”).

31. See *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1036-37 (N.D. Cal. 2025) (finding that Meta’s training of an LLM using downloaded books from “shadow libraries” is a fair use, but emphasizing that the “ruling does not stand for the proposition that Meta’s use of copyrighted materials to train its language models is lawful. . . [but] that these plaintiffs made the wrong arguments and failed to develop a record in support of the right one”); *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1023 (N.D. Cal. 2025) (finding that Anthropic’s copying of copyrighted books to train LLMs is a fair use). Note that neither of these cases addresses the possibility of an LLM outputting copies of a copyrighted work used in training.
32. See *infra* Parts II.B.-C.
33. See *infra* notes 106-08 and accompanying text; Crusey, *supra* note 18, at 501 (“Notably, district courts of the Ninth Circuit Court of Appeals have varied in their treatments of the identity requirement Application of the identity requirement is also unsettled in district courts beyond the Ninth Circuit.”); *Doe v. Github, Inc.*, No. 22-cv-06823, 2024 WL 4336532, at *2 (N.D. Cal. Sept. 27, 2024) (noting disagreement on the issue within the Northern District of California), *argued*, No. 24-7700 (9th Cir. Feb. 11, 2026).
34. 2024 WL 235217, at *8 (finding “that Section 1202(b) claims require that copies be ‘identical’”).
35. *Github*, 2024 WL 4336532, at *2-3 (N.D. Cal. Sept. 27, 2024) (granting interlocutory appeal on the issue of whether the DMCA imposes an identity requirement).

Though courts in favor of identity have focused on § 1202's requirement that the violating work be a copy of the original, the case law supporting identity on this ground is tenuous.³⁶ A defense of identity must instead be grounded in an honest acknowledgment that § 1202, under a similarity requirement, is technologically incompatible with the way LLMs are built. In fact, Congress expressly intended § 1202's scienter requirement to preclude liability for certain good-faith industry practices where CMI retention would be technologically infeasible.³⁷ As such, courts should read identity into § 1202's scienter requirement where, as with contemporary LLMs, further CMI retention is technologically infeasible. Doing so would allow the DMCA to complement traditional copyright, rather than supplant it.

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

The Digital Millennium Copyright Act was enacted in 1998 to update copyright law for the digital age.³⁸ Though it implements two World Intellectual Property Organization (WIPO) treaties, its conceptual origins trace back to a 1995 report published by the Clinton Administration.³⁹ That report recommended legislation prohibiting both the circumvention of copyright protection systems and the alteration of copyright management information⁴⁰—the two protections that later became the focus of § 1201 and § 1202 of the DMCA.⁴¹ The National Information Infrastructure Copyright Protection Act was drawn directly from the report.⁴² It stalled in Congress, largely due to an inability to reach agreement on “the contentious issue of the scope of liability of service providers for the infringing acts of their users.”⁴³ Congressional

36. See *infra* notes 115-19 and accompanying text.

37. See S. REP. NO. 105-190, at 37 (1998) (stating that technical infeasibility and financial hardship are grounds for avoiding § 1202(b) liability); see also *id.* (providing the transmission of excessively long credits in motion pictures and television broadcasts as an example of financial hardship).

38. See Digital Millennium Copyright Act, Pub. L. No. 105-304, 112 Stat. 2860 (1998) (codified at 17 U.S.C. §§ 512, 1201-05, 1301-32; 28 U.S.C. § 4001).

39. BRUCE A. LEHMAN, INFO. INFRASTRUCTURE TASK FORCE, INTELLECTUAL PROPERTY AND THE NATIONAL INFRASTRUCTURE: THE REPORT OF THE WORKING GROUP ON INTELLECTUAL PROPERTY RIGHTS (1995), <https://perma.cc/PE5Y-EKAN>.

40. *Id.* app. 1, at 5-7.

41. 17 U.S.C. §§ 1201-1202.

42. S. REP. NO. 105-190, at 3 (1998) (“On September 28, 1995, Chairman Hatch, with Senator Leahy, introduced the National Information Infrastructure (NII) Copyright Protection Act of 1995 (S. 1284), which embodied the legislative recommendations of the [Working Group on Intellectual Property Rights] White Paper.”).

43. *Id.* at 4.

momentum eventually resumed alongside efforts to implement the WIPO treaties, culminating in the passage of the DMCA.⁴⁴

An examination of the DMCA's legislative history reveals two important insights relevant to today's legal debates over LLMs. First, Congress expansively incorporated safe harbors and exemptions to accommodate technical challenges, shielding online service providers from traditional copyright infringement liability,⁴⁵ certain product manufacturers from § 1201 liability,⁴⁶ and broadcasters from § 1202 liability.⁴⁷ Second, the method of incorporating these protections suggests a path forward for LLM developers without modern legislative intervention. While all existing carveouts are now expressly noted by statute,⁴⁸ the legislative history demonstrates that the § 1202 broadcaster carveout was initially understood not as an explicit exception, but as an interpretation of the scienter requirement.⁴⁹ The statute may therefore accommodate further implicit carveouts.

44. *Id.* at 5 (describing how draft legislation implementing the World Intellectual Property Organization treaties “became the basis for Title I of the Digital Millennium Copyright Act”).

45. 17 U.S.C. § 512; *see* U.S. COPYRIGHT OFF., SECTION 512 OF TITLE 17: A REPORT OF THE REGISTER OF COPYRIGHTS 23-24 (2020), <https://perma.cc/GZ7E-8EQW>.

46. H.R. REP. NO. 105-551, pt. 1, at 10 (1998) (“The bill, as reported, presents a reasonable compromise by preventing only the manufacture or sale of devices that: (1) are ‘primarily designed’ to grant free, unauthorized access to copyrighted works; (2) have only limited commercially significant purpose or use other than to grant such free access; or (3) are intentionally marketed for use in granting such free access.”).

47. 17 U.S.C. § 1202(e).

48. *See id.* § 512 (exempting online service providers from infringement liability); *id.* § 1202(e) (exempting broadcasters from § 1202 liability); *see also id.* § 1201(c)(3) (echoing § 1201's accommodation of certain manufacturers by clarifying that they have no affirmative design obligation).

49. *See* H.R. REP. NO. 105-551, pt. 1, at 20-21 (1998) (“The prohibition in this subsection does not include ordinary and customary practices of broadcasters or inadvertent omission of credits from broadcasts of audiovisual works since, inter alia, such omissions do not entail knowing provision of false CMI with intent to induce, enable, facilitate or conceal a copyright infringement.”). Note that at that time, the draft text of § 1202 did not contain an express broadcaster exception. Though the express inclusion of carveouts in the text of the DMCA may suggest that Congress considered the excluded parties otherwise liable under the DMCA, *see* Eric F. Harbert, *In the Shadow of Mt. Olympus: Could a Revision of 17 U.S.C. §§ 1202-1204 Bring Them into Daylight?*, 13 UCLA ENT. L. REV. 133, 142 (2005) (“If such practices do not meet the minimum mental requirement for liability under section 1202, then there is no need for the exceptions discussed here because such ordinary practices would not trigger liability in the first place.”), legislative history suggests the intention was instead to resolve legal ambiguity, *see The Copyright Infringement Liability of Online and Internet Service Providers: Hearing on S. 1146 Before the S. Comm. on the Judiciary*, 105th Cong. 98 (statement of George Vradenburg III); *WIPO Copyright Treaties Implementation Act; and Online Copyright Liability Limitation Act: Hearing on H.R. 2281 and H.R. 2280 Before the H. Comm.*

footnote continued on next page

Challenges of the Digital Millennium.—In the 1990s, the explosive growth of the internet gave rise to a host of new challenges for parties throughout the copyright system. Digital technologies enabled the global distribution of millions of exact copies of copyrighted works within seconds, at minimal cost and with “virtually no degradation in quality.”⁵⁰ The invention of the MP3 file format demonstrates this phenomenon. Before MP3, “a person wishing to copy an original music recording . . . was limited to analog, rather than digital, recording technology.”⁵¹ When someone made, for example, a cassette tape, “each successive generation of copies [suffered] from an increasingly pronounced degradation in sound quality.”⁵² This decrease in quality disincentivized copying. In addition, media files were often too large for computers to store, and “downloading even a single song from the Internet took hours.”⁵³ MP3 and other “compression algorithms” enabled distribution over the internet by making media files smaller.⁵⁴ The invention of MP3 thus meant that sound recordings could “be copied virally and perpetually” at “near CD quality.”⁵⁵ As technology eliminated practical limitations on the scale and quality of both reproduction and distribution, copyright holders feared an exponential rise in infringement.⁵⁶

The Clinton Report and the National Information Infrastructure.—In September 1995, President Clinton’s Working Group on Intellectual Property Rights published *Intellectual Property and the National Information Infrastructure*, a detailed report assessing how emerging internet technologies would interact with existing intellectual property regimes.⁵⁷ The report proposed several legislative reforms, including extending legal safeguards to both technological protection measures and CMI, the two protections later enshrined in § 1201

on the Judiciary, 105th Cong. 87 (1997) [hereinafter *WIPO Treaties Hearing*] (statement of Roy Neel, U.S. Tel. Ass’n).

50. McCullough, *supra* note 15, at 90.

51. *Recording Indus. Ass’n of Am. v. Diamond Multimedia Sys., Inc.*, 180 F.3d 1072, 1073 (9th Cir. 1999).

52. *Id.*

53. *Id.*

54. *Id.* at 1073-74.

55. McCullough, *supra* note 15, at 81.

56. U.S. COPYRIGHT OFF., *supra* note 45, at 14 & n.46 (collecting examples) (citing NICHOLAS NEGROPONTE, *BEING DIGITAL* 58 (1995) (“In the digital world, not only the ease [of copying] is at issue, but also the fact that the digital copy is as perfect as the original and, with some fancy computing, even better.”); and Authors Guild, Inc., Comment Letter on Section 512 Study, at 3 (Apr. 1, 2016), <https://perma.cc/X37C-L3X7> (“Congress acknowledged the need for protection against digital piracy in the Internet age, due to the ease of copying and distributing perfect copies . . .”).

57. See LEHMAN, *supra* note 39, at 2.

and § 1202 of the DMCA.⁵⁸ Notably, the report differed from the later-enacted DMCA in its treatment of online service providers, concluding that “the best policy is to hold the service provider liable” so the free market might incentivize such providers to “make their subscribers more aware of copyright law.”⁵⁹

Protecting CMI was understood as a critical part of building a thriving, creative internet.⁶⁰ Legislation to implement the report’s recommendations was introduced in the National Information Infrastructure Copyright Protection Act of 1995, but the bill never passed.⁶¹ The Copyright Office had expressed concern about the breadth of the Act, “[g]iven the potential for substantial liability, and even felony criminal penalties.”⁶² Ultimately, legislative efforts stalled largely because lawmakers could not resolve disagreements about the scope of service provider liability for users’ infringing acts.⁶³

II. The DMCA in Practice: Liability Explodes

Section 1202(b) of the DMCA provides three distinct bases of liability related to CMI. Each of the three subsections is subject to the double scienter requirement—a construction fittingly named after the two scienter prongs a plaintiff must show to make out a claim for each violation. Section 1202(b)(1) prohibits the *intentional* removal or alteration of CMI; § 1202(b)(2) prohibits distribution of CMI with *knowledge* that CMI was removed or altered without authority; and § 1202(b)(3) prohibits the distribution of works or copies of

58. *Id.* app. 1, at 5-7; *see also* 17 U.S.C. §§ 1201-1202.

59. LEHMAN, *supra* note 39, at 117, 124.

60. *See NII Copyright Protection Act of 1995: Joint Hearing on H.R. 2441 and S. 1284 Before the Subcomm. on Cts. & Intell. Prop. of the H. Comm. on the Judiciary and the S. Comm. on the Judiciary*, 104th Cong. 52 (1995) [hereinafter *NII Copyright Protection Act Hearing*] (statement of Marybeth Peters, Register of Copyrights) (“If obtaining such information is difficult, people are more likely to forego the [National Information Infrastructure’s] benefits or resort to unauthorized uses.”); *id.* at 3 (statement of Sen. Orrin G. Hatch, Chairman, S. Comm. on the Judiciary) (framing digital copyright protections as necessary to promote creativity in the emerging online environment).

61. S. REP. NO. 105-190, at 4 (1998).

62. *NII Copyright Protection Act Hearing*, *supra* note 60, at 45, 52 (statement of Marybeth Peters, Register of Copyrights).

63. S. REP. NO. 105-190, at 4; *see also* NII Copyright Protection Act of 1995, H.R. 2441, 104th Cong. (1995); NII Copyright Protection Act of 1995, S. 1284, 104th Cong. (1995); Heidi Pearlman Salow, *Liability Immunity for Internet Service Providers: How Is It Working?*, 6 J. TECH. L. & POL’Y 31, 39-40 (2001) (“The Working Group considered the arguments from both sides, and concluded that it was premature to enact legislation that would reduce the liability of [internet service providers].”).

works with *knowledge* that CMI was removed or altered without authority.⁶⁴ The intent requirement associated with § 1202(b)(1) and the knowledge requirement associated with § 1202(b)(2) and (3) comprise the first prong of the double scienter requirement. All three subsections are additionally subject to the second prong, which requires actual or constructive *knowledge* that the action taken “will induce, enable, facilitate, or conceal an infringement.”⁶⁵

Courts may find copyright infringement without finding a § 1202 violation.⁶⁶ For example, an unauthorized reproduction that appropriately includes CMI infringes copyright but does not trigger § 1202 liability. Perhaps more controversially, the inverse is also true—works can violate § 1202 without infringing copyright.⁶⁷

Although § 1202 was intended to adapt copyright enforcement to the challenges of the digital age, it has evolved into a powerful standalone doctrine that threatens to expand liability far beyond traditional infringement. Consider the various differences between the standards for a § 1202(b) violation and copyright infringement. Direct infringement is a strict liability offense that explicitly requires copyrightable subject matter, follows a two-part test for proving liability, permits a fair use defense, and limits damages on a per-work basis.⁶⁸ By contrast, § 1202 comes with a loosely applied scienter requirement, no express copyrightability threshold, no recognized fair use defense, and damages that accrue on a per-violation basis.⁶⁹ In addition, whereas copyright registration is required to bring a copyright infringement claim, no such requirement applies to claims under § 1202.⁷⁰ The result is that

64. 17 U.S.C. § 1202.

65. *Id.* §§ 1201-1202.

66. *See Andersen v. Stability AI Ltd.*, 744 F. Supp. 3d 956, 969, 971 (N.D. Cal. 2024) (finding the plaintiffs’ allegations of infringement sufficient but dismissing their DMCA claims); *Crowley v. Jones*, 608 F. Supp. 3d 78, 86, 90 (S.D.N.Y. 2022) (same); *Free Speech Sys., LLC v. Menzel*, 390 F. Supp. 3d 1162, 1172, 1175 (N.D. Cal. 2019) (same).

67. *Schneider v. YouTube, LLC*, 649 F. Supp. 3d 872, 891 (N.D. Cal. 2023) (granting summary judgment for the defendant on infringement claims relating to twenty-seven works while denying summary judgment on the plaintiff’s DMCA claims). Note that *Doe v. GitHub* is an example of a case in which the plaintiffs allege violations of the DMCA without alleging copyright infringement. Second Amended Complaint, *supra* note 12, ¶¶ 204-05.

68. 17 U.S.C. § 1202; *see infra* Part II.

69. *See infra* Part II.

70. *SellPoolSuppliesOnline.com LLC v. Ugly Pools Ariz., Inc.*, 344 F. Supp. 3d 1075, 1081 (D. Ariz. 2018) (“The text of the DMCA does not limit the protection of CMI to registered works.”); *Shihab v. Complex Media, Inc.*, No. 21-cv-6425, 2022 WL 3544149, at *8 (S.D.N.Y. Aug. 17, 2022) (concluding that 17 U.S.C. § 412’s registration requirement does not apply to DMCA § 1202); *Crusey*, *supra* note 18, at 515 (“Unlike a copyright infringement claim under Section 501, a Section 1202 claim requires no prerequisite copyright registration.”).

§ 1202 has the potential to impose massive liability for removing CMI from minimally original works, for uses that would otherwise qualify as fair use, and for producing works that only loosely resemble the alleged source material.⁷¹ This Part examines the mechanisms of that expansion by comparing § 1202's scienter, copyrightability, liability threshold, defenses, and damages elements to those of traditional infringement doctrine.

A. Weakened Standard for Proving Copying

The standard for determining whether copying gives rise to liability in a copyright infringement claim consists of two parts: copying in fact and copying in law. Copying in fact asks whether copying actually happened, and copying in law determines whether that copying was legally significant.⁷² Unlike the infringement standard, the § 1202 standard for determining whether a copy is legally significant depends on the court. Some courts implement an identity standard, under which the allegedly violating work must be identical to the allegedly violated, original work.⁷³ Other courts instead implement a similarity standard, under which similarity, rather than identity, is enough to determine that a “copy” was made.⁷⁴

Infringement.—To show copying in fact, a plaintiff bringing an infringement claim must first show that the defendant actually copied the original work.⁷⁵ This inquiry exists to ensure plaintiffs and defendants did not independently create their respective works, a finding that would invalidate the infringement claim.⁷⁶

Because direct evidence of copying is rare, plaintiffs often use circumstantial evidence to show copying in fact, including some combination

71. Crusey, *supra* note 18, at 483 (“[D]amages have the potential to accumulate to massive amounts in cases involving removal of CMI en masse, which results in many consecutive individual violations of Section 1202(b). In many of the recent AI copyright lawsuits, CMI is alleged to have been removed from millions of works, leading to potential Section 1202 damage awards in the billions of dollars.”); *infra* notes 134-44 and accompanying text (explaining how thinly copyrighted works may form the basis of § 1202(b) claims).

72. See *infra* notes 75-89 and accompanying text.

73. See, e.g., cases cited *infra* note 102.

74. See, e.g., cases cited *infra* note 122.

75. RESTATEMENT OF COPYRIGHT § 7.03(a) (A.L.I., Tentative Draft No. 4, 2023) (“Proving copying in fact requires the plaintiff to prove some material was copied from the plaintiff’s copyrighted work . . .”).

76. *Id.* (“If the defendant proves that the allegedly infringing material was independently created rather than copied, directly or indirectly, from the plaintiff’s copyrighted work, there was no copying in fact from the plaintiff’s work and therefore no liability for copyright infringement.”).

of proof of the defendant's access to the plaintiff's work and the similarity between the works.⁷⁷

Striking similarity alone may also be enough to show copying.⁷⁸ This is a strict standard; the plaintiff must show identity or similarity "such that, considering all the circumstances, there is no reasonable explanation . . . other than copying."⁷⁹ Whether a work is strikingly similar overlaps somewhat with the "thinness" (or here, "thickness") of the work's copyright⁸⁰: "[T]he more inventive or expressive the duplicated content, the more likely the duplication resulted from copying."⁸¹

In addition to copying in fact, infringement claims must show "copying in law."⁸² This prong serves to determine whether an established act of copying legally rises to the level of infringement.⁸³ In particular, copying in law ensures that only copyrightable material can be infringed⁸⁴ and that the copying of this material is "substantial."⁸⁵

To that end, in the Ninth Circuit, for example, the substantial similarity test involves filtering for protectable components in the plaintiff's work

77. Lydia Pallas Loren & Anthony R. Reese, *Proving Infringement: Burdens of Proof in Copyright Infringement Litigation*, 23 LEWIS & CLARK L. REV. 621, 641-43 (2019); RESTATEMENT OF COPYRIGHT, *supra* note 75, § 7.03.

78. RESTATEMENT OF COPYRIGHT, *supra* note 75, § 7.03.

79. *Id.*

80. *See infra* notes 87-89 and accompanying text.

81. Robert F. Helfing, *Substantial Similarity and Junk Science: Reconstructing the Test of Copyright Infringement*, 30 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 735, 763-64 (2020).

82. Mark Edward Blankenship, Jr., *Dead Frogs, Dissected Jokes & Thin Copyright: Analyzing Copyrightable Elements and Legal Protection of Stand-Up Comedy*, 2 FLA. ENT. & SPORTS L. REV. 117, 127 (2023).

83. RESTATEMENT OF COPYRIGHT, *supra* note 75, § 7.04 ("Proving improper appropriation requires the plaintiff to prove that sufficient protected expression was copied from the copyrighted work such that there is a substantial similarity between the alleged infringing use and the copyrighted work.").

84. *See Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 363 (1991) ("As a constitutional matter, copyright protects only those constituent elements of a work that possess more than a *de minimis* quantum of creativity."); Blankenship, *supra* note 82, at 127 ("[Copying in law] considers whether the defendant copied a copyrightable expression in such a way that liability should follow."); 4 NIMMER & NIMMER, *supra* note 22, § 13D.06 ("[T]he substantial similarity step of the infringement analysis focuses only on the protected elements of plaintiff's work that defendant copied." (emphasis omitted)).

85. Alan Latman, "Probative Similarity" as Proof of Copying: Toward Dispelling Some Myths in Copyright Infringement, 90 COLUM. L. REV. 1187, 1189 (1990) ("[N]ot only must [the defendant] copy protected material, but also such protected material must be 'substantial.'" (first quoting *R.L. Polk & Co. v. Musser*, 196 F.2d 1020, 1021 (3d Cir. 1952) (per curiam); then quoting *Mathews Conveyer Co. v. Palmer-Bee Co.*, 135 F.2d 73, 84-85 (6th Cir. 1943); and then quoting *Wilson v. Haber Bros.*, 275 F. 346, 347 (2d Cir. 1921))).

before conducting an objective comparison of the two works and then determining whether the “total concept and feel” of the two works is similar.⁸⁶ While the normal infringement standard requires substantial similarity, “when a work is entitled to ‘thin’ protection, courts typically require ‘virtual identity’” instead.⁸⁷ This means the validity or strength of a work’s copyright protection moderates the level of similarity required for a finding of substantial similarity. That is, where copyright is at its thinnest, as with compilations of factual material,⁸⁸ “liability hinges on entire duplication.”⁸⁹

Various doctrines govern whether material is copyrightable. The idea-expression dichotomy, for example, permits protection only for a work’s expressive elements—not its underlying ideas.⁹⁰ For instance, a book describing a “particular bookkeeping system” is itself copyrightable, but the method of bookkeeping it describes is not.⁹¹ Similarly, facts are not protected,⁹² though the “selection and arrangement” of facts in a compilation may be.⁹³ The merger doctrine holds that if there are few ways of expressing an idea, that expression “merges” with the idea and is unprotectable.⁹⁴ For example, a jewel-encrusted pin shaped like a bee is unprotected because there are limited ways to depict that idea.⁹⁵ Finally, works in the public domain are

86. *Rentmeester v. Nike, Inc.*, 883 F.3d 1111, 1118 (9th Cir. 2018) (quoting *Cavalier v. Random House, Inc.*, 297 F.3d 815, 822 (9th Cir. 2002)), *overruled on other grounds*, *Skidmore v. Led Zeppelin*, 952 F.3d 1051 (9th Cir. 2020) (en banc).

87. *Advanta-STAR Auto. Rsch. Corp. of Am. v. Search Optics, LLC*, 672 F. Supp. 3d 1035, 1049 (S.D. Cal. 2023) (quoting *Eta-Hokin v. Skey Spirits Inc.*, 323 F.3d 763, 766 (9th Cir. 2003); *Kaseberg v. Conaco, LLC*, 260 F. Supp. 3d 1229, 1244 (S.D. Cal. 2017); and *Rassamni v. Fresno Auto Spa, Inc.*, 365 F. Supp. 3d 1039, 1045, 1047 (E.D. Cal. 2019)).

88. *Feist*, 499 U.S. at 349 (1991) (“This inevitably means that the copyright in a factual compilation is thin.”).

89. 4 NIMMER & NIMMER, *supra* note 22, § 13D.32 (adding that multiple circuits agree with this approach, and that “virtual identity is not the same as absolute identity”).

90. 1 NIMMER & NIMMER, *supra* note 22, § 2.03.

91. Edward Samuels, *The Idea-Expression Dichotomy in Copyright Law*, 56 TENN. L. REV. 321, 326-27 (1989).

92. *Eldred v. Ashcroft*, 537 U.S. 186, 217 (2003) (“[C]opyright gives the holder no monopoly on any knowledge. A reader of an author’s writing may make full use of any fact or idea she acquires from her reading.”).

93. *Feist*, 499 U.S. at 350 (“As applied to a factual compilation, assuming the absence of original written expression, only the compiler’s selection and arrangement may be protected; the raw facts may be copied at will.”).

94. William E. Greenspan, *Applications of the Idea/Expression Merger Doctrine in Copyright Infringement Cases*, BUS. L. REV., Spring 1993, at 1, 2.

95. *Herbert Rosenthal Jewelry Corp. v. Kalpakian*, 446 F.2d 738, 742 (9th Cir. 1971). Instead of finding the expression entirely unprotectable, some courts instead afford it “thin” copyright protection. See J.H. Reichman, *Electronic Information Tools—The Outer Edge of World Intellectual Property Law*, 17 U. DAYTON L. REV. 797, 818 (1992) (“[P]recedents suggest that copyright law should never afford borderline factual and functional
footnote continued on next page”).

not protected by copyright.⁹⁶

Section 1202.—Section 1202’s framework does not include an analysis of copying in fact or copying in law. The provision does, however, require that a violating work is a “copy” of the violated work.⁹⁷ This requirement comes from § 1202(c)’s definition of “copyright management information” as various types of “information conveyed in connection with copies . . . of a work.”⁹⁸ In other words, a work creating § 1202 liability cannot be a *new* work that is unrelated to the allegedly violated work. To ensure this standard is met, courts have developed two different standards that serve as proxies for whether the work in question is a copy of the original, allegedly violated work.⁹⁹

works more than ‘thin’ protection against wholesale appropriation of surface expression.”).

96. Faith O. Majekolagbe, *Public Domain and Access to Knowledge*, 31 J. INTELL. PROP. L. 1, 2 (2024).

97. 17 U.S.C. § 1202(c) (defining “copyright management information” as “information conveyed in connection with copies”). Note that the provision again refers to copies in § 1202(b)(3), which specifies “copies of works” in its prohibition of distribution, import for distribution, and public performance of works.

98. *Id.* § 1202(c).

99. One understanding is that courts derive their identity and similarity standards from the requirement that a work be a “copy.” 4 NIMMER & NIMMER, *supra* note 22, § 12A.10(C)(1) (finding that § 1202(b)(3)’s reference to copies “serves as the statutory hook to prevent nonsensical claims that plaintiff has a right to complain about elimination of its copyright management information when defendant propounds an entirely dissimilar work”); *Real World Media LLC v. Daily Caller, Inc.*, 744 F. Supp. 3d 24, 40 (D.D.C. 2024) (opining—while declining to take a position—that the identity requirement is “grounded in the DMCA’s definition of CMI as information ‘conveyed in connection with copies . . . of a work’” (alteration in original) (emphasis omitted) (quoting 17 U.S.C. § 1202(c)); *N.Y. Times Co. v. Microsoft Corp.*, 777 F. Supp. 3d 283, 317 (S.D.N.Y. 2025) (“While the DMCA does not define ‘copies of works,’ an abundance of case law establishes that in ‘cases where claims of removal of CMI have been held viable, the underlying work has been substantially or entirely reproduced.” (quoting *Fischer v. Forrest*, 286 F. Supp. 3d 590, 609 (S.D.N.Y. 2018), *aff’d*, 968 F.3d 216 (2d Cir. 2020))). However, some courts instead justify these standards using the statutory prohibition against “remov[ing] . . . copyright management information.” § 1202(b)(1). Such courts state that merely omitting CMI is not a violation of § 1202(b), whereas removing CMI is. They require identity or similarity to show that the allegedly violating work is not a recreation of the original work from which CMI was omitted, but rather the exact same work without CMI. *E.g.*, *Dolls Kill, Inc. v. Zoetop Bus. Co.*, No. 222-cv-01463, 2022 WL 16961477, at *4 (C.D. Cal. Aug. 25, 2022) (“The differences between the parties’ products undercut any inference that Defendants removed or altered Plaintiff’s CMI. To find otherwise would be to speculate that by merely omitting CMI, Defendants must have necessarily removed it.”); *Andersen v. Stability AI Ltd.*, 744 F. Supp. 3d 956, 971 (N.D. Cal. 2024) (claiming that lack of identity pleads a plaintiff out of a § 1202(b) claim “because failing to affix CMI to a ‘different work’ is not ‘removal’ under Section 1202”); *Kipp Flores Architects, LLC v. Pradera SFR, LLC*, No. SA-21-CV-00673, 2022 WL 1105751, at *3 (W.D. Tex. Apr. 13, 2022)

footnote continued on next page

Some courts hold that a § 1202 violation requires the challenged work to be identical to the original work (an “identity requirement”).¹⁰⁰ Other courts disagree, holding instead that some threshold of similarity is sufficient (a “similarity requirement”).¹⁰¹ There is a split among courts and commentators regarding which of these tests is the proper standard for determining whether a work is a “copy.”¹⁰² The Ninth Circuit will soon decide which standard is applicable to § 1202 claims within the Circuit—the similarity standard or the identity standard.¹⁰³

The courts implementing an identity requirement consist mostly of district courts within the Ninth Circuit.¹⁰⁴ These courts have held that a challenged work must be an identical copy of the original to violate § 1202.¹⁰⁵

(“Removal’ of CMI from a copyrighted work is not the same as the failure to add CMI to a nonidentical rendition or a derivative of the protected work.”). *But see* Doe v. GitHub, Inc., 672 F. Supp. 3d 837, 857 (N.D. Cal. 2023) (stating that the “semantic distinction” between passive noninclusion and active removal of CMI is “not meaningful”).

100. *See, e.g.*, cases cited *infra* note 104.

101. *See* cases cited *infra* note 122.

102. *Real World Media LLC v. Daily Caller, Inc.*, 744 F. Supp. 3d 24, 40 (D.D.C. 2024) (stating that there is a “nascent district-court split” in which “[s]ome courts have held that no DMCA violation exists where an allegedly infringing work is not an ‘identical copy’ of the original but rather is a derivative work or recreates aspects of the original work’s protected expression” while “other courts have held that a DMCA claim may lie under such circumstances” (quoting *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-1931, 2020 WL 5991503, at *6 (C.D. Cal. Aug. 14, 2020)); *Doe v. Github, Inc.*, No. 22-cv-06823, 2024 WL 4336532, at *2 (N.D. Cal. Sept. 27, 2024) (referring to disagreement on the identity requirement as a “split in authority”). *Compare* 4 NIMMER & NIMMER, *supra* note 22, § 12A.10(C)(1) (“[T]o be actionable, the removal must be from a work that is at least substantially similar to the copyrighted work in question. . . . But it goes too far to conclude, as did *dictum* in one case, that ‘no DMCA violation exists where the works are not identical.’” (quoting *O’Neal v. Sideshow, Inc.*, 583 F. Supp. 3d 1282, 1287 (C.D. Cal. 2022))), *with* PAUL GOLDSTEIN, GOLDSTEIN ON COPYRIGHT § 7.18 (Wolters Kluwer 3d ed. Supp. 2026) (“As a rule, the works must be identical for there to be a violation.”).

103. *See* Doe v. GitHub, Inc., No. 22-cv-06823, 2024 WL 235217, at *8 (N.D. Cal. Jan. 22, 2024), *argued*, No. 24-7700 (9th Cir. Feb. 11, 2026).

104. *See, e.g.*, *GitHub*, 2024 WL 235217, at *8 (dismissing a § 1202(b) claim for lack of identity); *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.” (citation omitted)); *O’Neal*, 583 F. Supp. 3d at 1287 (“[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*, 2020 WL 5991503, at *6 (“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.”).

105. *See* cases cited *supra* note 104.

However, even individual courts have been inconsistent in whether they apply the identity requirement. 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;¹⁰⁶ one found in contrast that a § 1202(b) claim *survived* a motion to dismiss where the challenged work was “a derivative of [the plaintiff’s] copyrighted source code” (and therefore presumably not identical);¹⁰⁷ and one found that the “split in authority” on this issue contributed to its inability to dismiss on the basis of identity.¹⁰⁸

Even those courts that have seemingly aligned around the existence of an identity standard have failed to agree on how to apply such a standard. Some cases have suggested the relevance of a challenged work’s status as a derivative work in determining whether it is identical to the original,¹⁰⁹ with one court explaining that derivative works cannot violate the DMCA because they are distinct from the original.¹¹⁰ Another court required perfect identity, dismissing a § 1202(b) claim for lack of identity where plaintiffs alleged a “functional[ly] equivalent,” “nearly verbatim reproduction” with only “a few ‘cosmetic’ differences in word choice.”¹¹¹ 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.”¹¹² Different still,

106. *Tremblay v. OpenAI, Inc.*, 716 F. Supp. 3d 772, 779-80 (N.D. Cal. 2024) (“Under the plain language of the statute, liability requires distributing the original ‘works’ or ‘copies of [the] 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.”).

107. *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 [the plaintiff’s] copyrighted source code. For now, however, the CMI is sufficiently pleaded to survive a motion to dismiss.”).

108. *Beijing Meishe Network Tech. Co. v. TikTok Inc.*, No. 23-cv-06012, 2024 WL 3522196, at *9 (N.D. Cal. July 23, 2024).

109. *E.g., Frost-Tsuji Architects v. Highway Inn, Inc.*, No. 13-00496, 2015 WL 263556, at *4 (D. Haw. Jan. 21, 2015) (finding that a derivative use of the plaintiff’s work, rather than a use of the original work, cannot uphold a CMI removal claim under § 1202(b)).

110. *Crowley v. Jones*, 608 F. Supp. 3d 78, 90 (S.D.N.Y. 2022).

111. *Doe v. GitHub, Inc.*, No. 22-cv-06823, 2024 WL 235217, at *2, *9 (N.D. Cal. Jan. 22, 2024) (quoting First Amended Complaint ¶¶ 103-04, 10, *GitHub*, 2024 WL 235217 (Nos. 22-cv-06823 & 22-cv-07074), 2023 WL 4843613, ECF No. 135).

112. *O’Neal v. Sideshow, Inc.*, 583 F. Supp. 3d 1282, 1287 (C.D. Cal. 2022).

another court determined lack of identity by visually analyzing side-by-side images, providing no further details about the analysis.¹¹³ 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.”¹¹⁴ 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.

Courts that do not follow the identity requirement have criticized these decisions as unsupported.¹¹⁵ Indeed, the cases that adopt an identity standard are largely self-referencing in their reasoning; all cite back to *Kirk Kara Corp. v. Western Stone & Metal Corp.*,¹¹⁶ a single case decided by the District Court for the Central District of California.¹¹⁷ There, the court became the “first court to affirmatively impose an ‘identity’ requirement” for § 1202(b),¹¹⁸ while itself relying on cases that do not expressly require identity.¹¹⁹ At least one of the cases cited imposes an arguably *stricter*

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

114. *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999), *aff’d in part, rev’d in part*, 336 F.3d 811 (9th Cir. 2003).

115. *See, e.g.*, *ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d 411, 427 (S.D. Tex. 2023) (“Although the court in *Kirk Kara* held the DMCA requires identical copies, the caselaw it cited does not support its holding. The court relied on *Fischer*, which did not employ an identical copies requirement Likewise, the *Kirk Kara* court cited two other cases, but neither mentioned nor employed an identical copies requirement.” (citations omitted) (citing *Fischer v. Forrest*, 286 F. Supp. 3d 590, 609-10 (S.D.N.Y. 2018)); *New Parent World, LLC v. True to Life Prods., Inc.*, No. CV-23-08089, 2024 WL 4277865, at *2 (D. Ariz. Sept. 24, 2024) (“*Kirk Kara* states that ‘no DMCA violation exists where the works are not identical,’ but it relies on three cases that do not support this position.” (citation omitted)). This criticism has also been levied in a leading treatise. 4 NIMMER & NIMMER, *supra* note 22, § 12A.10(C)(1) n.150.7 (“Turning to *Kirk Kara*, that unpublished decision cited two reported cases. Neither relied on the *identity* standard.” (citation omitted)).

116. No. CV 20-1931, 2020 WL 5991503 (C.D. Cal. Aug. 14, 2020).

117. *Advanta-STAR Auto. Rsch. Corp. of Am. v. Search Optics, LLC*, 672 F. Supp. 3d 1035, 1057 (S.D. Cal. 2023) (citing *Kirk Kara*, 2020 WL 5991503, at *6); *O’Neal*, 583 F. Supp. 3d at 1287 (same); *GitHub*, 2024 WL 235217, at *9 (same); *Tremblay v. OpenAI, Inc.*, 716 F. Supp. 3d 772, 779-80 (N.D. Cal. 2024) (same); *see also Andersen v. Stability AI Ltd.*, 744 F. Supp. 3d 956, 971 & n.11 (N.D. Cal. 2024) (relying on *GitHub*, 2024 WL 235217, at *8-9, which itself cites *Kirk Kara*).

118. *Crusey*, *supra* note 18, at 500.

119. *Kirk Kara*, 2020 WL 5991503, at *6 (citing *Kelly*, 77 F. Supp. 2d at 1122 (requiring CMI to be removed from “a plaintiff’s product or original work” for § 1202(b)(1) to apply); *Fischer v. Forrest*, 286 F. Supp. 3d 590, 609 (S.D.N.Y. 2018) (noting that the *only* similarity between the original and allegedly violating work consists of “four discrete phrases” and finding the rule in past § 1202(b)(1) cases to require that works are “substantially or entirely reproduced”), *aff’d*, 968 F.3d 216 (2d Cir. 2020); *Frost-Tsuji*

footnote continued on next page

standard than identity, requiring CMI to be removed from “a plaintiff’s product or original work.”¹²⁰ That standard seems to inherently encompass identity: A work that is merely similar to the original has been modified in some way and is thus not original. Another of the three cases *Kirk Kara* cites suggests at the very least that identity is *evidence* of CMI removal from an “original” work.¹²¹ *Kirk Kara*’s establishment of an identity requirement thus has more support than its critics suggest.

On the other side of the split, courts and commentators have held that something less than identity is sufficient to show that CMI was removed from a “copy.”¹²² In one case, a district court held that “superficial alterations” to a work—alterations that in at least some courts on the other side of the split would lead to a dismissal for lack of identity—could not justify the defendants’ motion to dismiss.¹²³ The court held that “there is no justification, based on congressional intent, before the Court for adding the extra term of ‘identical’ to the DMCA’s plain language.”¹²⁴

Though some courts do not specify a standard beyond a rejection of identity,¹²⁵ others have specifically stated that the correct standard is

Architects v. Highway Inn, Inc., No. 13-00496, 2015 WL 263556, at *3 (D. Haw. Jan. 21, 2015) (suggesting that CMI must be removed from an original work)).

120. *Kelly*, 77 F. Supp. 2d at 1122.

121. *Frost-Tsuji Architects*, 2015 WL 263556, at *3 (suggesting that the court could have inferred removal of CMI had the defendant’s drawing been identical to the plaintiff’s).

122. *See, e.g.*, Oracle Int’l Corp. v. Rimini St., Inc., No. 14-cv-01699, 2023 WL 4706127, at *82 (D. Nev. July 24, 2023) (rejecting the defendant’s argument “that a work that removes copyright management information must be an exact copy of the original work”), *vacated in part*, 123 F.4th 986 (9th Cir. 2024); *ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d 411, 417 (S.D. Tex. 2023) (“Specifically, the Court is persuaded by the reasoning of courts that have found that the DMCA may properly apply even when the allegedly infringing work is not identical to the original.”); *New Parent World, LLC v. True to Life Prods., Inc.*, No. CV-23-08089, 2024 WL 4277865, at *2 (D. Ariz. Sept. 24, 2024) (“Other courts have held that an infringing work need not be an identical copy to violate the DMCA. The Court finds these cases persuasive.” (citations omitted)); 4 NIMMER & NIMMER, *supra* note 22, § 12A.10(C)(1) (“[T]o be actionable, the removal must be from a work that is at least substantially similar to the copyrighted work in question But it goes too far to conclude, as did *dictum* in one case, that ‘no DMCA violation exists where the works are not identical.’” (quoting *O’Neal*, 583 F. Supp. 3d at 1287)).

123. *ADR Int’l*, 667 F. Supp. 3d at 428-29.

124. *Id.* at 430.

125. *See, e.g.*, *Real World Media LLC v. Daily Caller, Inc.*, 744 F. Supp. 3d 24, 40-41 (D.D.C. 2024) (“Ultimately, the Court agrees with a leading treatise that the arguments requiring perfect identity under these circumstances ‘fail[]] to withstand scrutiny.’ . . . Hence, the Court rejects Daily Caller’s categorical argument that exact copies of portions of a video can never support a DMCA claim.” (citing 4 NIMMER & NIMMER, *supra* note 22, § 12A.10(C)(1)); *ADR Int’l*, 667 F. Supp. 3d at 425 (“The DMCA is not limited to CMI conveyed in connection with identical copies of a work.”).

whether the challenged and original works are “substantially similar.”¹²⁶ One Ninth Circuit case abided by a slightly different standard: “striking similarity.”¹²⁷ The court imported this standard directly from copyright infringement doctrine, illustrating courts’ occasional confusion between copyright infringement and DMCA § 1202 claims.¹²⁸ In that case, striking similarity was found where allegedly infringing photographs contained “alterations, such as tint effects and added logos” that only appeared in the original source.¹²⁹ Few courts have followed this approach.¹³⁰ Others have defined their standard by reference to a work’s derivative nature rather than its similarity or identity; some courts explicitly state that derivative works are not exempt from § 1202 liability.¹³¹ But no general rule exists: Just like the

126. See, e.g., *Enter. Tech. Holdings v. Noveon Sys., Inc.*, No. 05-CV-2236, 2008 WL 11338356, at *7 (S.D. Cal. July 29, 2008) (finding the defendant violated the DMCA by altering CMI from “transportation management software” he programmed “that was ‘substantially similar’” to the original software); 4 NIMMER & NIMMER, *supra* note 22, § 12A.10(C)(1) (noting that DMCA liability requires the defendant’s work to be “at least substantially similar” to the protected work).

127. *Friedman v. Live Nation Merch., Inc.*, 833 F.3d 1180, 1188 (9th Cir. 2016) (quoting *Baxter v. MCA, Inc.*, 812 F.2d 421, 423 (9th Cir. 1987)).

128. *Friedman* cites *Baxter* for its standard on striking similarity. *Friedman*, 833 F.3d at 1188. But *Baxter* does not involve any DMCA claims; it merely discusses a “striking similarity” standard as a means of establishing copying in a copyright infringement claim. 812 F.2d at 423, 434 n.2. The *Friedman* court neither acknowledges that it is borrowing from infringement doctrine nor distinguishes the standard from § 1202’s statutory inquiry. See 833 F.3d at 1188. If the court were consciously importing an infringement standard, one would expect at least some discussion of its applicability to the DMCA.

129. *Friedman*, 833 F.3d at 1188.

130. At least two district courts within the Ninth Circuit have acknowledged the Ninth Circuit’s striking similarity standard. *Aardwolf Indus., LLC v. Abaco Machs. USA, Inc.*, No. CV 16-1968, 2017 WL 10350547, at *11 (C.D. Cal. Nov. 13, 2017); *Oracle Int’l Corp. v. Rimini St., Inc.*, No. 14-cv-01699, 2023 WL 4706127, at *82 (D. Nev. July 24, 2023), *vacated in part*, 123 F.4th 986 (9th Cir. 2024). But many more district courts within the Ninth Circuit have declined to do so. See *supra* notes 103-04 (listing district court cases within the Ninth Circuit that have followed the identity standard instead).

131. See, e.g., *Huffman v. Activision Publ’g, Inc.*, No. 19-cv-00050, 2020 WL 8678493, at *12 (E.D. Tex. Dec. 14, 2020) (“Nothing in the statute exempts derivative works or a defendant’s own work.”); *GC2 Inc. v. Int’l Game Tech.*, 391 F. Supp. 3d 828, 843-44 (N.D. Ill. 2019) (rejecting the “broad proposition that derivative or collaborative works are categorically excluded from protection under the DMCA’s provision for removal of copyright management information,” though never explicitly endorsing either side of the court split on identity). But see 4 NIMMER & NIMMER, *supra* note 22, § 12A.10(C)(1) (conceding that “an unauthorized film may infringe the copyright in the novel on which it is based, but crediting the film to someone other than the novelist does not seem to rise to the level of false copyright management information” while claiming that the proper standard is substantial similarity (footnote omitted)).

courts on the identity side of the split, most courts advocating a similarity standard have not specified the application of that standard.¹³²

Copyrightability.—Notably, unlike the standard for proving copying under infringement, neither standard for determining whether a work is a “copy” under § 1202 considers the copyrightability of the underlying subject matter.¹³³ Unlike the infringement standard, where the copying in fact inquiry determines whether a work was actually copied and the copying in law inquiry determines if that copying led to substantial similarity sufficient for legal action, § 1202 uses either the identity or the similarity standard as a proxy for both of those inquiries. But in effect, this collapses the inquiries: The similarity and identity standards require some measure of similarity between the works in question, and once that similarity is established, it is seemingly assumed to be legally significant.¹³⁴ Whereas the infringement

132. See, e.g., *Real World Media LLC v. Daily Caller, Inc.*, 744 F. Supp. 3d 24, 40-41 (D.D.C. 2024) (rejecting identity without explaining how similarity could be applied); *Oracle*, 2023 WL 4706127, at *82 (same); *ADR Int'l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d 411, 429 (S.D. Tex. 2023) (same); *New Parent World, LLC v. True To Life Prods., Inc.*, No. CV-23-08089, 2024 WL 4277865, at *2 (D. Ariz. Sept. 24, 2024) (same). Where courts reject identity, alignment with some similarity standard can be reasonably assumed. It is unclear what the alternative might be to identity or similarity, besides doing away with § 1202 in its entirety.

133. Note, however, that 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. *Schneider v. YouTube, LLC*, 674 F. Supp. 3d 704, 723-24 (N.D. Cal. 2023). Another court focused on ownership rather than the extent of the work's ability to be copyrighted. See *Sanborn Libr. LLC v. ERIS Info., Inc.*, No. 19-CV-2049, 2024 WL 1744630, at *20 (S.D.N.Y. Mar. 25, 2024) (finding “issues of material fact as to which copyrights [plaintiff] owns and may enforce”). 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, the “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.*, No. 23 CV 3424, 2025 WL 933951, at *8 (N.D. Ill. Mar. 27, 2025) (quoting *Design Basics, LLC v. Signature Constr., Inc.*, 994 F.3d 879, 886 (7th Cir. 2021)). In support of this contention, the court cited a case that is entirely about copyright infringement and does not mention § 1202 once. *Id.* (citing *Design Basics*, 994 F.3d at 886). The closest a court has come to properly acknowledging the need for a copyrightability analysis was the statement in *Raw Story Media, Inc. v. OpenAI, Inc.* that the plaintiffs “nowhere alleged that the information in their articles is copyrighted.” 756 F. Supp. 3d 1, 7 (S.D.N.Y. 2024). However, there the court was referring to LLMs’ “synthes[is] of the relevant information” from its training data, which 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.

134. See *Oracle*, 2023 WL 4706127, at *82 (applying the substantial similarity standard to find for the plaintiff on DMCA claims without analyzing the copyrightability of the

footnote continued on next page

analysis devotes its copying in law prong to a determination of the underlying work's copyrightability, asking whether the similarities established in the copying in fact analysis are across protectable elements, the § 1202 analysis seems to cross its fingers and hope for the best.

While copyrightability may seem more germane to infringement, it is equally necessary for § 1202 liability. It is true that, in stark contrast to the thicket of doctrines governing copyrightable subject matter for copyright infringement claims, § 1202 does not contain an express requirement that violated works be copyrightable. However, commentators and courts have held that works must pass some threshold of nonzero copyrightability to comprise valid § 1202 subject matter.¹³⁵ Common sense implies the same: In addition to intent (for § 1202(b)(1)) or knowledge (for § 1202(b)(2) and (3)), the double scienter requirement demands that a § 1202 violator know or have reason to know their activity will “induce, enable, facilitate, or conceal an infringement.”¹³⁶ It follows that the work must comprise copyrightable subject matter; otherwise, it would not be capable of being infringed, and this aspect of the scienter requirement would be impossible to satisfy.¹³⁷

Despite the logical necessity that the underlying work be copyrightable, courts do not distinguish between a work's copyrighted and uncopyrighted elements when determining the similarity or identity of two works under § 1202.¹³⁸ Works could thus theoretically be found similar or identical despite

underlying work, despite stating that the work is “copyrighted”); *Friedman*, 833 F.3d at 1188 (reversing the district court's grant of summary judgment for the defendant under a striking similarity standard, without analyzing the copyrightability of the underlying work).

135. See, e.g., GOLDSTEIN, *supra* note 102, § 7.18 (“Although the provision nowhere expressly requires that the subject matter in issue be under copyright, the requirement in section 1202(a) of an intention ‘to induce, enable, facilitate, or conceal infringement,’ like the introductory reference in 1202(b) to ‘the authority of the copyright owner,’ and in the concluding clause to ‘an infringement of any right under this title,’ implies that copyright exists in the subject matter.” (quoting U.S.C. § 1202(a)-(b))); *Schneider v. YouTube, LLC*, 674 F. Supp. 3d 704, 718 (N.D. Cal. 2023) (finding that “[o]wnership of a copyright is an essential predicate for [CMI] claims”).

136. 17 U.S.C. § 1202(b); see *infra* Part II.D (discussing the double scienter requirement).

137. See GOLDSTEIN, *supra* note 102, § 7.18.

138. See, e.g., *Fischer v. Forrest*, 286 F. Supp. 3d 590, 609-10, 609 n.9 (S.D.N.Y. 2018) (failing to analyze the copyrightability of any of the four works from which CMI may have been removed, though coming close in stating that one work “could, in theory, constitute a ‘work’ under the DMCA”), *aff'd*, 968 F.3d 216 (2d Cir. 2020); *Stevens v. Corelogic, Inc.*, 899 F.3d 666 (9th Cir. 2018) (omitting a discussion of copyrightability); *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 172 (2d Cir. 2020) (stating incidentally that § 1202(b)(3) applies to “copyrighted material” while finding that the district court—which omitted a copyrightability analysis—correctly applied the DMCA); *Enter. Tech. Holdings v. Noveon Sys., Inc.*, No. 05-CV-2236, 2008 WL 11338356, at *16 (S.D. Cal., July 29, 2008) (finding for the plaintiff where the defendant admitted to modifying a work's CMI

footnote continued on next page

sharing few copyrightable elements. This broadens the scope of § 1202 immensely compared to infringement, where the validity of the subject matter under copyright law is considered multiple times throughout the analysis.¹³⁹

Take, for example, an original work consisting of a list of the author’s ten favorite movies. If a copy of this work were assessed under copyright infringement, the copying in law analysis would consider which aspects of the list are protectable. Because the list is essentially a compilation of facts, the analysis would determine that the copyright, overall, is quite thin; it is likely that only the selection and arrangement of the listed elements would be protected.¹⁴⁰ And because the copyright is thin, it is likely that the applicable standard would be closer to identity than to substantial similarity.¹⁴¹

This is exactly the approach a leading treatise suggests for assessing this fact pattern under § 1202: “[W]hen the work in question has only thin creativity, the applicable standard may approach identity”¹⁴² In analyzing the list of ten favorite movies for § 1202 liability, the treatise claims that “[t]o be substantially similar might require copying, in order, at least nine, and possibly all ten” list items.¹⁴³ But despite the treatise’s optimism in applying “traditional copyright doctrine” to the DMCA context,¹⁴⁴ no courts have taken its proposed approach when analyzing a § 1202 claim. Thinly copyrighted works remain capable of forming the basis for successful § 1202 claims even where they might fall to infringement’s more exacting standard.

without determining the copyrightability of the underlying work); sources cited *supra* note 134 (listing § 1202 cases in which the copyrightability of the allegedly violated work was not analyzed). *But see Schneider*, 674 F. Supp. 3d at 723-24 (stating that class certification for a § 1202(b) claim requires common proof of copyright ownership). However, the dispute centers around the existence of *class-wide* proof of ownership—not the extent and substance of the ownership itself.

139. *See Lexmark Int’l, Inc. v. Static Control Components, Inc.*, 387 F.3d 522, 538 (6th Cir. 2004) (clarifying that “[b]oth prongs of the infringement test”—that is, both the validity of the copyright and the substantial similarity of the allegedly infringing and allegedly infringed work—“consider ‘copyrightability’” (quoting *Brown Bag Software v. Symantec Corp.*, 960 F.2d 1465, 1476 (9th Cir. 1992))).

140. *See Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 349 (1991) (“This inevitably means that the copyright in a factual compilation is thin.”).

141. *See Advanta-STAR Auto. Rsch. Corp. of Am. v. Search Optics, LLC*, 672 F. Supp. 3d 1035, 1049 (S.D. Cal. 2023).

142. 4 NIMMER & NIMMER, *supra* note 22, § 12A.10(C)(1).

143. *Id.* § 12A.10(C)(1) n.150.9.

144. *Id.* § 12A.10(C)(1) n.150.8.

B. No Fair Use Defense

A claim of copyright infringement can be countered by a prima facie rebuttal or an affirmative defense.¹⁴⁵ Prima facie rebuttals “can best be understood as assertions by the defendant that the plaintiff has not successfully met its burden of proof on some element or elements of its infringement claim.”¹⁴⁶ For example, a defendant may argue that the plaintiff’s copyright is not valid or that plaintiff is not the owner of the copyright. Affirmative defenses under infringement include the usual suspects: Defendants can argue that their use falls outside the statute of limitations or within the scope of a valid license.

These affirmative defenses seem to be allowed under § 1202 as well.¹⁴⁷ However, fair use—which scholars have often referred to as the “most important defense” against infringement¹⁴⁸—appears to be considered by many courts as an invalid defense for § 1202 claims.¹⁴⁹ Indeed, a purely textual

145. Loren & Reese, *supra* note 77, at 627-28 (categorizing “affirmative defenses” and prima facie factual and legal rebuttals as “defense arguments,” though the latter “are not technically affirmative defenses”).

146. *Id.* at 648.

147. FurnitureDealer.Net, Inc v. Amazon.com, Inc., No. 18-232, 2022 WL 891473, at *21 (D. Minn. Mar. 25, 2022) (citation omitted) (“The statute of limitations for a DMCA claim is three years. The Complaint was filed on January 26, 2018, therefore any alleged removal of CMI that occurred prior to January 26, 2015 falls outside the statute of limitations period unless that period is tolled.”); Intercept Media, Inc. v. OpenAI, Inc., 767 F. Supp. 3d 18, 33 (S.D.N.Y. 2025) (considering—though ultimately rejecting for unrelated reasons—the defendant’s statute of limitations defense); Schneider v. YouTube, LLC, 674 F. Supp. 3d 704, 718 (N.D. Cal. 2023) (stating that “the existence of a license . . . is an essential defense” against claims relating to CMI); Loeb-Defever v. Strategic Constr., Ltd., No. 20-CV-1981, 2022 WL 22877320, at *12 (S.D. Tex. Jan. 27, 2022) (finding the existence of a license to negate the scienter element of a possible DMCA claim).

148. Brian L. Frye, *Aesthetic Nondiscrimination & Fair Use*, 3 BELMONT L. REV. 29, 30 (2016) (“The most important defense to copyright infringement is the fair use doctrine”); Clark D. Asay, Essay, *Transformative Use in Software*, 70 STAN. L. REV. ONLINE 9, 9 (2017) (“Fair use is copyright law’s most important defense against claims of copyright infringement.”) (citing MARSHALL A. LEAFFER, UNDERSTANDING COPYRIGHT LAW 495 (6th ed. 2014) (claiming fair use is “by far the most important defense to an action for copyright infringement”).

149. Susuk Lim, *A Survey of the DMCA’s Copyright Management Information Protections: The DMCA’s CMI Landscape After All Headline News and McClatchey*, 6 WASH. J.L. TECH. & ARTS 297, 307 (2011) (“At the most textual end of the spectrum, § 1202 makes no exception for fair use; removal or alteration of copyright information is categorically prohibited ‘without the authority of the copyright owner or law.’” (quoting David Johnson, *Court Split Widens over Whether DMCA Rules Against Removal of Copyright Management Information Apply Only to Automatic, Computerized Copyright Management Systems*, DAVID JOHNSON’S DIGIT. MEDIA LAW. BLOG (Mar. 25, 2009), <https://perma.cc/4CBF-BDPF>)); see cases cited *infra* note 151 (applying a fair use analysis to infringement claims but not to § 1202 claims).

reading of the provision makes no mention of fair use.¹⁵⁰ It has rarely been asserted as a defense to a § 1202 claim.¹⁵¹ On the other hand, § 1202's prohibition of the removal or alteration of CMI "without the authority of . . . the law"¹⁵² has been suggested by some scholars as a possible source of a fair use requirement.¹⁵³

Arguments in support of fair use's application to § 1202 are not entirely persuasive. One such argument maintains that fair use precludes fulfillment of § 1202's double scienter requirement.¹⁵⁴ This requirement states that an accused § 1202(b) violator must know or have "reasonable grounds to know" that their removal, alteration, or distribution of CMI "will induce, enable, facilitate, or conceal an infringement."¹⁵⁵ The idea is that removing or altering the CMI cannot "induce, enable, facilitate, or conceal an infringement" because fair use is definitionally not infringement.¹⁵⁶

150. Lim, *supra* note 149, at 307.

151. See *Murphy v. Millennium Radio Grp. LLC*, No. 08-1743, 2015 WL 419884, at *5 (D.N.J. Jan. 30, 2015) (finding that the application of fair use does not preclude the necessary scienter for a § 1202 claim); *Free Speech Sys., LLC v. Menzel*, 390 F. Supp. 3d 1162, 1173-75 (N.D. Cal. 2019) (analyzing the application of fair use to an infringement claim and dismissing a § 1202 claim for unrelated reasons without a fair use analysis); *Monroe v. NorthStar Source Grp., LLC*, No. 23-cv-06220, 2025 WL 193902, at *5-8, 11-12 (S.D.N.Y. July 15, 2025) (analyzing fair use for an infringement claim but not for a § 1202 claim); *Shihab v. Source Digit., Inc.*, No. 23cv7266, 2024 WL 3461351, at *3-9 (S.D.N.Y. July 18, 2024) (same); *Reiner v. Canale*, 301 F. Supp. 3d 727, 741-44, 749 (E.D. Mich. 2018) (same); *Bell v. Milwaukee Bd. of Sch. Dirs.*, No. 22-C-0227, 2022 WL 18276966, at *4-8 (E.D. Wis. Dec. 21, 2022) (same); *Bain v. Film Indep., Inc.*, No. CV 18-4126, 2020 WL 5491314, at *3-6 (C.D. Cal. Aug. 6, 2020) (same), *aff'd*, No. 20-55948, 2022 WL 17592422 (9th Cir. Dec. 13, 2022); *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1118-23 (C.D. Cal. 1999) (same). A handful of cases have allowed the defense. See *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417, 2025 WL 1786418, at *1 (N.D. Cal. June 27, 2025) ("The plaintiffs' DMCA claim fails because . . . [the defendants'] copying must be deemed fair use"); *Kennedy v. Gish, Sherwood & Friends, Inc.*, 143 F. Supp. 3d 898, 914 (E.D. Mo. 2015) (finding that fair use precludes a § 1202 violation). One of those decisions has been criticized for basing its fair use conclusion on a misunderstanding of § 1201 and a misapplication of § 1201 to § 1202. See 4 NIMMER & NIMMER, *supra* note 22, § 12A.10(B)(2)(b) n.96.1 ("The court's conclusion would therefore appear to be in error." (citing *Kennedy*, 143 F. Supp. 3d at 914)).

152. 26 U.S.C. § 1202(b).

153. See 4 NIMMER & NIMMER, *supra* note 22, § 12A.10(B)(2)(b); Eric F. Harbert, *In the Shadow of Mt. Olympus*, 13 UCLA ENT. L. REV. 133, 145 (2005) ("Section 1202(b) also allows removal or alteration of CMI by authority of 'the law.' . . . Failure to clarify this section of the statute could have a profound chilling effect upon the fair use of copyrighted works."); Crusey, *supra* note 18, at 511.

154. See *infra* Part II.D.

155. 26 U.S.C. § 1202(b); see *infra* Part II.D (discussing the double scienter requirement).

156. See *infra* Part II.D; Lim, *supra* note 149, at 307 ("[B]ecause fair use is fundamentally not infringement, the removal of CMI, whether or not intentional, cannot lead to actual infringement. Thus, one required element of a § 1202 claim would be unfulfilled.");

footnote continued on next page

However, this argument is misguided: Though a fair use is not itself infringement, it could certainly enable infringement. For example, assume a chapter from a textbook that is copied and disseminated to students for use in an educational setting qualifies as a fair use.¹⁵⁷ If the work's CMI were removed, a scenario could easily arise in which the work became disqualified from fair use by its use in a different context—for example, if a student were to distribute the work in a commercial setting. This could still be an infringement even though the original use was deemed fair.¹⁵⁸

Placing fair use applications of copyrighted material beyond the scope of § 1202 contradicts congressional intent in the same manner. The legislative purpose of § 1202 was to prevent CMI removal so second-order users would not unknowingly—and thus undeterred by copyright incentives—violate copyright law.¹⁵⁹ As such, it is counterproductive to rule out § 1202 when the first-order use is fair use, because § 1202 is concerned with the undesirability of the second, third, and fourth-order uses. Consider, as another example, how a fair use defense might affect the ability of open-source developers to ensure that derivative works remain open to the public.¹⁶⁰ Software is modular, meaning that sections of code are often reused in other projects.¹⁶¹ Since the early days of the internet, licenses have promoted the open dissemination of code through provisions that allow developers to copy modules so long as they make their products open source as well.¹⁶² But under fair use, a first-order

Kadrey, 2025 WL 1786418, at *1 (“[T]he Copyright Act provides that anything that is a fair use ‘is not an infringement of copyright.’ So because [the defendant’s] copying was not an infringement, its removal of CMI could not have furthered an act of infringement.” (citation omitted) (quoting 17 U.S.C. § 107)).

157. See 17 U.S.C. § 107 (stating that one factor in the fair use analysis is “the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes”).

158. See Harbert, *supra* note 153, at 137 (providing another example of a fair use facilitating infringement).

159. See *supra* notes 14-17 and accompanying text; *NII Copyright Protection Act Hearing*, *supra* note 60, at 51-52.

160. Brian W. Carver, Note, *Share and Share Alike: Understanding and Enforcing Open Source and Free Software Licenses*, 20 BERKELEY TECH. L.J. 443, 455-56, 470 (2005) (evaluating the scope and enforceability of “copyleft” licenses).

161. See Gregory D. Schwartz, Essay, *When Disciplines Disagree: The Admissibility of Computer-Generated Forensic Evidence in the Criminal Justice System*, 72 UCLA L. REV. DISCOURSE 174, 185 (2024) (“[A] modern software product is code that runs other code that runs other code.”).

162. See, e.g., V.K. Unni, *Fifty Years of Open Source Movement: An Analysis Through the Prism of Copyright Law*, 40 S. ILL. U. L.J. 271, 280-82 (2015) (discussing the Mozilla Public License, which requires licensees to make available the source code for any modifications to the original software); Peter S. Menell, *Economic Analysis of Network Effects and Intellectual Property*, 34 BERKELEY TECH. L.J. 219, 261-63 (2019) (discussing the more stringent

footnote continued on next page

user engaged in scholarship or research may not be obligated to comply with such licenses. Such a user can publish code that is substantially similar to the original work while ignoring requirements to include CML.¹⁶³ This, combined with the internet’s capacity for near-instant, lossless proliferation,¹⁶⁴ could then induce countless subsequent users to effectively copy the original work while unknowingly violating the original license.¹⁶⁵ A § 1202 fair use defense would thus compromise a key mechanism for sustaining the programming community’s historic and vibrant open-source culture.¹⁶⁶

Another scholar has raised the logistical difficulties of importing a fair use requirement into § 1202. First, “one cannot simply insert a reference” to fair use into § 1202 “because such protections would only be triggered if there were an infringement in the first place.”¹⁶⁷ There is little basis to assume fair use applies to § 1202 where the provision has not been adapted to trigger without the occurrence of an infringement.¹⁶⁸ Second, it would “[be unwise] to integrate an independent ‘fair use’ type structure into section 1202” as it is “an area that Congress has clearly left to the courts.”¹⁶⁹ Plaintiffs making § 1202 claims would certainly agree with these arguments, having predictably argued that § 1202 defendants cannot bring a fair use defense.¹⁷⁰

“copyleft” General Public License, which grants licensees broad permissions and requires them to license any resulting software under the same terms).

163. *Cf., e.g., Mozilla Public License Version 1.1*, MOZILLA, § 3.3, <https://perma.cc/Y2KM-NHBU> (archived Dec. 31, 2025) (“You must include a prominent statement that the Modification is derived, directly or indirectly, from Original Code provided by the Initial Developer and including the name of the Initial Developer in (a) the Source Code, and (b) in any notice in an Executable version or related documentation in which You describe the origin or ownership of the Covered Code.”).

164. *See supra* notes 15-16 and accompanying text.

165. *Cf., e.g., Mozilla Public License, supra* note 163, § 3.2 (requiring source code availability).

166. Unni, *supra* note 162, at 298 (“[T]he open source licensing movement in general has undoubtedly contributed to the progress of the new information society and has laid a strong foundation for economic growth.”); Josh Lerner & Jean Tirole, *The Scope of Open Source Licensing* 11 n.15 (Nat’l Bureau of Econ. Rsch., Working Paper No. 9363, 2002), <https://perma.cc/CFG4-7BX7> (describing how copyleft licenses encourage more copyleft licenses).

167. Harbert, *supra* note 153, at 145.

168. *See* 17 U.S.C. § 107 (stating that “the fair use of a copyrighted work . . . is not an infringement of copyright”).

169. Harbert, *supra* note 153, at 145.

170. Second Amended Complaint, *supra* note 12, at 22 (“The Fair Use affirmative defense is only applicable to Section 501 copyright infringement. It is not a defense to violations of the DMCA It cannot be used to avoid liability here.”); Plaintiffs’ Reply to Motion for Summary Judgment and Opposition to Meta’s Motion for Partial Summary Judgment at 39 n.30, *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026 (N.D. Cal. 2025) (No. 23-cv-03417), 2025 WL 1166301, ECF No. 537 (“Meta’s argument that Plaintiffs’ DMCA claim fails if there is a finding of fair use lacks merit.”).

Amid this ambiguity, even stakeholders who believe that fair use *should* be an available defense to § 1202(b) claims acknowledge that “[i]t is not entirely certain” if fair use is available.¹⁷¹ A leading treatise summarized the confusion nicely, noting that fair use’s application to § 1202 is but one unanswered question of the “wealth of dangling questions” Congress left for courts to resolve in applying § 1202.¹⁷² Perhaps due to this confusion, § 1202 defendants generally have not asserted fair use as a defense.¹⁷³ In practice, then, infringement defendants seem to have access to a significant defense that is effectively unavailable to § 1202 defendants, expanding the reach of § 1202 to fair uses.

C. Heightened Statutory Damages

Just as with infringement,¹⁷⁴ under the DMCA, “[p]laintiffs are given a choice between receiving actual or statutory damages.”¹⁷⁵ Certainly, damages for infringement have the potential to be very high. But statutory damages assessed under infringement are mitigated by two factors: (1) fair use may render them inapplicable, and (2) they attach per *work*, rather than per *violation*, meaning that each work can only produce one statutory damage award, regardless of how many times its copyright is violated.¹⁷⁶ Statutory damages under § 1202 have no such mitigating factors. As discussed in Part II.B, fair use is effectively an unavailable defense to § 1202 defendants. And under § 1202, statutory damages attach per *violation*, meaning that using one work to infringe multiple times generates multiple statutory damage awards.

Statutory damages for infringement start at \$750 per work infringed.¹⁷⁷ The minimum statutory damages award under § 1202(b) is \$2,500 per

171. Dave Hansen, *Updates on AI Copyright Law and Policy: Section 1202 of the DMCA, Doe v. Github, and the UK Copyright and AI Consultation*, AUTHORS ALL. (Mar. 7, 2025), <https://perma.cc/TK28-FLYP>.

172. 4 NIMMER & NIMMER, *supra* note 22, § 12A.10.

173. *See* cases cited *supra* note 151 (considering fair use for infringement claims but not for § 1202 claims).

174. Chesley, *supra* note 26, at 24 (“[U]nder a traditional copyright infringement claim an infringer is liable for either the copyright owner’s actual damages, plus the infringer’s additional profits, or statutory damages. The copyright owner may decide, at any time before a final judgment, which of these he wishes to recover.” (footnote omitted)).

175. *Id.* at 22.

176. *See id.* at 4-5 (“The large penalties that have begun to be levied in DMCA cases are rarer under traditional copyright doctrine, because there are several mechanisms by which the Copyright Act prevents such an outcome. First, a single copyrighted work is considered in its entirety, and damages are assessed for the work as a whole Secondly, statutory awards are limited to one per work, not one per act of infringement.” (footnote omitted)).

177. 17 U.S.C. § 504.

violation—more than three times the minimum for infringement.¹⁷⁸ Though the statutory maximum is considerably less than that of infringement (\$25,000 vs. \$150,000 for willful infringement),¹⁷⁹ this difference is diminished by the cumulative effect of § 1202(b)'s per-violation damages scheme.

Despite little statutory guidance, courts generally seem to agree that the number of violations for purposes of § 1202 statutory damages is determined by “the number of discrete [violative] acts” taken by the defendant.¹⁸⁰ For example, one court counted only one violation where more than 1,000 people received a work in violation of § 1202 because the distribution of the work involved only a single act.¹⁸¹ One scholar questioned what the outcome would have been in this case had the defendant distributed each of the works in separate emails, concluding that “[u]nder the court’s definition of a ‘violation,’ it seems that each email sent would be a separate violation for damages purposes.”¹⁸² A different court addressed this scenario, finding that “providing the same computer program at different times, to different hospitals, would constitute distinct violations of § 1202.”¹⁸³

Another case, *GC2, Inc. v. International Game Technology*,¹⁸⁴ underscores the difference between infringement’s “per work” damages scheme and § 1202’s “per violation” damages scheme. There, the jury found 696 DMCA violations for the defendants’ uploading of copyrighted game material to user-accessible servers.¹⁸⁵ The jury found that “each reupload constituted three separate violations because three separate versions of the [copyrighted] works found to have had removed or altered copyright management information were uploaded.”¹⁸⁶ The court awarded the lowest possible statutory damages: \$2,500 per violation.¹⁸⁷ Even so, the § 1202 statutory damages totaled \$1.74 million.¹⁸⁸ This award would have been far less under infringement’s “per work” rule; because only two games were infringed, damages would have been assessed for

178. *Id.* § 1203.

179. *Id.* §§ 504, 1203.

180. Chesley, *supra* note 26, at 22.

181. *Id.* at 21 (citing *McClatchey v. Associated Press*, No. 05-cv-145, 2007 WL 1630261, at *6 (W.D. Pa. June 4, 2007)).

182. *Id.* at 23.

183. *Id.* (citing *Goldman v. Healthcare Mgmt. Sys., Inc.*, 559 F. Supp. 2d 853 (W.D. Mich. 2008)).

184. 391 F. Supp. 3d 828 (N.D. Ill. 2019).

185. *GC2*, 391 F. Supp. 3d at 838.

186. *Id.* at 850.

187. *Id.* at 838.

188. *Id.*

only two distinct works.¹⁸⁹ Unsurprisingly, the plaintiff opted for actual instead of statutory damages for its infringement claims.¹⁹⁰

Courts have varied in their willingness to assess § 1202 statutory damages on the high end of the available range. One court stated that “courts generally award statutory damages on the lower end of the damages spectrum”¹⁹¹ but cited a court that, just three years later, found the *maximum* award per violation appropriate in a case where determining actual damages was “more than ‘difficult’” and where the defendant’s conduct was “willful and even duplicitous.”¹⁹² These are factors courts routinely consider in determining the proper amount of statutory damages per violation.¹⁹³ One issue with this framework is that all § 1202(b) violations are to some degree “willful”—after all, a successful claim necessitates that the scienter requirement has been met. This redundancy diminishes the factor’s value and risks allowing courts to invoke willfulness as a broad justification for imposing high statutory damages—to the detriment of some defendants.¹⁹⁴

D. The Intent Requirement Fails to Cabin Liability

Direct copyright infringement is based on a strict liability rule.¹⁹⁵ This means that no knowledge of infringement is necessary for liability, and innocent infringement is not a defense.¹⁹⁶ The Copyright Act thus “casts a wide

189. See Chesley, *supra* note 26, at 5 (“As a result of this limitation, an infringer selling thousands of copyrighted books would be liable for one statutory damage award per title.”).

190. GC2, 391 F. Supp. 3d at 838.

191. Wright v. Miah, No. 22-CV-4132, 2023 WL 6219435, at *13 (E.D.N.Y. Sept. 7, 2023).

192. Bus. Casual Holdings v. TV-Novosti, No. 21-CV-2007, 2023 WL 1809707, at *10-11 (S.D.N.Y. Feb. 8, 2023).

193. Miller v. Netventure24 LLC, No. 19-CV-7172, 2021 WL 3934262, at *8 (S.D.N.Y. Aug. 6, 2021) (listing “the difficulty of proving actual damages, the circumstances of the violation, whether [d]efendants violated the DMCA intentionally or innocently, and deterrence” as factors courts consider in determining statutory damages amounts (alteration in original) (quoting Agence France Presse v. Morel, No. 10-cv-2730, 2014 WL 3963124, at *10 (S.D.N.Y. Aug. 13, 2014))).

194. See *infra* Part III.

195. A. Samuel Oddi, *Contributory Copyright Infringement: The Tort and Technological Tensions*, 64 NOTRE DAME L. REV. 47, 52 (1989).

196. GOLDSTEIN, *supra* note 102, § 7.0.1 (“Strict liability is the rule in copyright cases, and the defendant who copies protected expression from a copyrighted work will be liable regardless of his innocence. . . . Under the rule of strict liability, a bookstore that sells an unauthorized copy of a book from among the thousands on its shelves will be liable for infringing the public distribution right even though it had no knowledge that the copy infringed.”).

net that ensnares infringers of all different stripes,¹⁹⁷ making it a powerful tool for copyright owners. In contrast, § 1202(b) has not one but two scienter requirements intended to cabin its reach, a construction that has fittingly been referred to as a “double scienter requirement.”¹⁹⁸ But in practice, courts have sometimes interpreted these requirements loosely. Indeed, the Southern District of New York commented on the Second Circuit’s “leniency when evaluating scienter on a motion to dismiss.”¹⁹⁹

The first scienter requirement states that the violative action itself must be intentional (for § 1202(b)(1)) or done with knowledge (for § 1202(b)(2) and (3)). For example, a defendant accused of violating § 1202(b)(1) for removing or altering CMI must have “intentionally” done so, and a defendant accused of violating § 1202(b)(3) for distributing a copy of a work without CMI must “know[] that the copyright management information has been removed or altered without authority of the copyright owner or the law.”²⁰⁰ Though this prong is ostensibly one half of a twofold scienter requirement, some courts have conflated it with the act of removing or altering CMI under § 1202(b)(1) such that the scienter requirement effectively consists only of the second prong for those claims. In one case, the Southern District of New York found the plaintiff’s complaint sufficient to allege knowledge of CMI removal where the defendant “allegedly removed the CMI itself without authorization from [the plaintiff].”²⁰¹ This is the only explanation provided by the court in support of its finding; it is essentially a restatement of the prong itself.²⁰² The same result arose in another case, where the same court reasoned that there was no material dispute as to intent because the defendant accessed the plaintiff’s work from Twitter, added it to a compilation, and publicly shared the compilation without attribution to the plaintiff.²⁰³ Similar to the prior case, these showings of intent effectively amount to the prohibited act itself. The court seems to find that removing CMI—or even just using a tool that removes CMI—is enough to

197. Dane S. Ciolino & Erin A. Donelon, *Questioning Strict Liability in Copyright*, 54 RUTGERS L. REV. 351, 351 (2002).

198. *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 171 (2d Cir. 2020).

199. *N.Y. Times Co. v. Microsoft Corp.*, 777 F. Supp. 3d 283, 315 (S.D.N.Y. 2025).

200. 17 U.S.C. § 1202(b)(1)-(2).

201. *Shihab v. Complex Media, Inc.*, No. 21-cv-6425, 2022 WL 3544149, at *5 (S.D.N.Y. Aug. 17, 2022).

202. 17 U.S.C. § 1202(b) (“No person shall, without the authority of the copyright owner or the law—(1) intentionally remove or alter any copyright management information[.]”); *Shihab*, 2022 WL 3544149, at *5 (“[T]he Court concludes that the [complaint] adequately alleges that [the defendant] knew that the CMI had been removed or altered without the authority of [plaintiff] or the law—as it allegedly removed the CMI itself without authorization from [plaintiff].”).

203. *Gwinn v. City of Chicago*, No. 23 CV 1823, 2025 WL 964089, at *7 (N.D. Ill. Mar. 31, 2025).

show *intent* to remove CMI. The District Court for the District of Columbia and Ninth Circuit district courts have held similarly.²⁰⁴

The second scienter requirement states that a defendant under § 1202 must have “actual or constructive knowledge” that their violative action “will induce, enable, facilitate, or conceal an infringement.”²⁰⁵ The idea here is that removing CMI from a work obscures the work’s allowed uses, such that future consumers of the work may unintentionally infringe. To show the required scienter under this prong, courts including the Eleventh and Ninth Circuits have looked for “a past ‘pattern of conduct’ or ‘modus operandi’” showing “the defendant was aware [of] or had reasonable grounds to be aware of the probable future impact of its actions.”²⁰⁶ The Second Circuit has similarly found that a defendant should have known his alteration of CMI concealed infringement where he “understood from his training and experience that he was required to get permission” to use the works in question.²⁰⁷ According to the Ninth Circuit, a “general possibility” of future infringement is not enough—if it were, the knowledge requirement would be superfluous, as there is a general possibility of future infringement “whenever CMI is removed.”²⁰⁸ On the other hand, plaintiffs do not have to show that future infringement is certain—only that it is likely.²⁰⁹ A “wide range of evidence” may be used to prove scienter.²¹⁰

Some courts within the above circuits have found a sufficient pattern of conduct with little evidence. In one Eastern District of New York case, the plaintiff submitted a DMCA takedown notice to Etsy, accusing the defendant of selling items on Etsy’s platform that featured the plaintiff’s artwork without the proper CMI.²¹¹ Granting default judgment, the court held that the

204. Compare *Pierce v. Lifezette, Inc.*, No. 20-0693, 2021 WL 2557241, at *5 (D.D.C. June 2, 2021) (finding mere allegations of intentional CMI removal, without more, sufficient to “plausibly [allege] a violation” of the first double scienter prong), and *Bain v. Film Indep., Inc.*, No. CV 18-4126, 2018 WL 6930766, at *1, *5 (C.D. Cal. Aug. 9, 2018) (finding removal of a watermark sufficient to plead a § 1202 claim), with *Logan v. Meta Platforms, Inc.*, 636 F. Supp. 3d 1052, 1064 (N.D. Cal. 2022) (“Unlike editing a plaintiff’s watermark out of a photo, automatically omitting CMI by embedding a photo out of the full context of the webpage where the CMI is found cannot itself plead intentionality as required by the DMCA.”).

205. *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 171 (2d Cir. 2020) (quoting 17 U.S.C. § 1202(b)(3)).

206. *Victor Elias Photography, LLC v. Ice Portal, Inc.*, 43 F.4th 1313, 1320 (11th Cir. 2022) (alteration in original) (quoting *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 674 (9th Cir. 2018)).

207. *Mango*, 970 F.3d at 173.

208. *Stevens*, 899 F.3d at 673-74.

209. *Id.* at 675.

210. *Schneider v. YouTube, LLC*, 674 F. Supp. 3d 704, 725 (N.D. Cal. 2023) (citing *Stevens*, 899 F.3d at 675; and *Victor Elias*, 43 F.4th at 1323).

211. *Wright v. Miah*, No. 22-CV-4132, 2023 WL 6219435, at *1 (E.D.N.Y. Sept. 7, 2023).

defendant's submission of a counternotice to Etsy disputing liability plausibly demonstrated the knowledge required under this prong.²¹² The Southern District of California found on a motion to dismiss that the knowledge requirement for § 1202(b)(2), the section of the provision disallowing distribution of removed or altered CMI, was "clearly . . . satisfied" where the defendant alleged "two prior instances in which [the defendant] admitted that it 'reproduced and distributed to its customers some of [the plaintiff's] copyrighted content' without authorization."²¹³ The court found this "pattern of conduct" sufficient to show "that the defendant was aware or had reasonable grounds to be aware of the probable future impact of its actions."²¹⁴ The Northern District of California required even less evidence to find scienter sufficient to deny a defendant's motion to dismiss, finding in one case that the "pattern of conduct" standard in fact only applies at the summary judgment stage and that earlier stages require only general allegations of scienter.²¹⁵ Lastly, though anomalous,²¹⁶ at least one district court within the circuits above has not conformed at all to the circuit courts' requirement of a "past 'pattern of conduct'"²¹⁷: The Southern District of New York found the plaintiff to "easily satisfy the second double scienter prong for concealment of a copyright infringement" at the motion to dismiss stage simply because the defendant posted photographs to social media "without [the plaintiff's]

212. *Id.* at *9 ("As required, the Court also finds it plausible that defendant knew or had reason to know that the removal of plaintiff's CMI would induce the infringement of a right protected under this section of the statute, given defendant's counterntices to Etsy.").

213. *Advanta-STAR Auto. Rsch. Corp. of Am. v. Search Optics, LLC*, 672 F. Supp. 3d 1035, 1057-58 (S.D. Cal. 2023). The motion to dismiss was, however, granted on other grounds. *Id.*

214. *Id.* at 1057.

215. *Izmo, Inc. v. Roadster, Inc.*, No. 18-cv-06092, 2019 WL 13210561, at *4 (N.D. Cal. Mar. 26, 2019) ("Critically, however, *Stevens* concerned a plaintiff's burden of proof at summary judgment, not at the pleading stage. *Stevens* does not contradict Rule 9(b)'s instruction that 'intent, knowledge, and other conditions of a person's mind may be alleged generally.'" (citation omitted) (quoting FED. R. CIV. P. 9(b))).

216. Courts within the Eleventh, Ninth, and Second Circuits typically conform to the "pattern of conduct" analysis for the second scienter prong. *See, e.g., Affordable Aerial Photography, Inc. v. Witkowski*, No. 24-cv-80423, 2025 WL 1755033, at *5 (S.D. Fla. Feb. 27, 2025) ("Accordingly, that evidence does not support Affordable Aerial's proposition that Beach Town had a 'pattern' of removing CMI from copyrighted works."); *Harrington v. Pinterest, Inc.*, No. 20-cv-05290, 2022 WL 4348460, at *5 (N.D. Cal. Sept. 19, 2022) (citing the Ninth Circuit's standard in dismissing the plaintiff's DMCA claims); *Schneider v. YouTube, LLC*, 674 F. Supp. 3d 704, 724 (N.D. Cal. 2023) (citing the Ninth Circuit's standard in denying the plaintiff's class certification request for a DMCA claim).

217. *Victor Elias Photography, LLC v. Ice Portal, Inc.*, 43 F.4th 1313, 1320 (11th Cir. 2022) (quoting *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 674 (9th Cir. 2018)).

authorization.”²¹⁸ Similarly, courts in other circuits have at times required little evidence in finding the second prong.²¹⁹

In addition to the risk of future third-party infringement, a defendant’s actual or constructive knowledge of their *own* infringement is enough to satisfy § 1202(b)’s knowledge requirement.²²⁰ This makes satisfaction of the knowledge requirement relatively simple where a defendant is accused of violating § 1202(b)(2) by distributing a work knowing CMI was removed. For these claims, as long as the defendant’s act of distribution is unauthorized and does not qualify for an infringement defense, the defendant may satisfy the knowledge requirement simply by engaging in the act subjecting them to § 1202 liability in the first place.²²¹

A low scienter bar at the motion to dismiss stage may alone create uncertainty and thus deter beneficial uses of copyrighted works. A low bar at the summary judgment stage further chills uses of copyrighted works for fear of expensive, drawn-out litigation. And though some courts have of course suggested higher standards for showing scienter, even those that appear the strictest are often applied loosely in practice.²²²

Infringement’s strict liability rule remains a harsher standard for defendants than § 1202’s scienter requirement, despite the unpredictability with which the latter is applied. And while § 1202’s scienter requirement “is intended to counterbalance [the statute’s] otherwise broad definition of

218. *Shihab v. Complex Media, Inc.*, No. 21-cv-6425, 2022 WL 3544149, at *5 (S.D.N.Y. Aug. 17, 2022).

219. *See McClatchey v. Associated Press*, No. 05-cv-145, 2007 WL 776103, at *6 (W.D. Pa. Mar. 9, 2007) (suggesting that the act of removing CMI from a work before distributing the work to subscribers may itself show “requisite intent to induce, enable, facilitate or conceal infringement” at summary judgment); *Millennium Funding, Inc. v. Priv. Internet Access, Inc.*, No. 21-cv-01261, 2022 WL 7560395, at *18 (D. Colo. Oct. 13, 2022) (emphasizing the low evidentiary burden for scienter at the pleading stage).

220. *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 172 (2d Cir. 2020) (“[A]n infringement’ is not limited to the infringing acts of third parties. The plain meaning of the statutory language also encompasses an infringement committed by the defendant himself. This includes the knowing, unauthorized infringement that serves as the basis of establishing the first scienter element of Section 1202(b).” (quoting 17 U.S.C. § 1202(b))).

221. *See, e.g., Furnituredealer.net, Inc v. Amazon.com, Inc*, No. 18-232, 2022 WL 891473, at *25 (D. Minn. Mar. 25, 2022) (“If Defendants are liable for copyright infringement, it would be entirely reasonable for a jury to find Defendants had the requisite scienter under § 1202(b) because removal of the CMI would conceal their own infringement.”).

222. *Compare Stevens v. Corelogic, Inc.*, 899 F.3d 666, 674 (9th Cir. 2018) (establishing that “specific allegations” and “an affirmative showing, such as by demonstrating a past ‘pattern of conduct,’” are required to find scienter (quoting *United States v. Todd*, 627 F.3d 329, 334 (9th Cir. 2010))), *with Advanta-STAR Auto. Rsch. Corp. of Am. v. Search Optics, LLC*, 672 F. Supp. 3d 1035, 1058 (S.D. Cal. 2023) (finding only two prior instances of unauthorized distribution of copyrighted content sufficient to demonstrate a “pattern of conduct”).

CMI”²²³ and “protect innocent CMI removers,”²²⁴ in practice it does not strike this balance.

III. Applying the DMCA to LLMs

The importance of § 1202’s scope has dramatically increased with the rise of LLMs. *Doe v. GitHub*, the Ninth Circuit interlocutory appeal set to decide the identity-similarity court split discussed in Part II.A, centers around LLMs’ DMCA liability.²²⁵ Similar cases have also cropped up in recent years, including *Intercept v. OpenAI*,²²⁶ *Raw Story Media v. OpenAI*,²²⁷ and *New York Times v. Microsoft*²²⁸ in the Southern District of New York, and *Tremblay v. OpenAI*²²⁹ in the Northern District of California. Though much of the focus on LLMs’ legality has been on copyright infringement, with dozens of cases brought on that issue,²³⁰ these DMCA cases have the potential to preempt their outcome by fating LLM providers to exorbitant damages without the possibility of a fair use defense.

LLM providers are particularly disadvantaged by the state of the law, as it is difficult for them to avoid violating § 1202 with their current technological capacity.²³¹ Indeed, all of the elements that make § 1202 claims broader than infringement claims—that is, intent, the standard for proving copying,

223. Keenan, *supra* note 24, at 421.

224. Marissa Truskowski, Note, *What’s in a Name: Understanding Copyright Management Information*, 49 FLA. ST. U. L. REV. 995, 1021-22 (2022).

225. See Appellants’ Opening Brief at 1, *Doe v. Github, Inc.*, No. 24-7700 (9th Cir. argued Feb. 11, 2026), ECF No. 18.1.

226. *Intercept Media, Inc. v. OpenAI, Inc.*, 767 F. Supp. 3d 18 (S.D.N.Y. 2025).

227. *Raw Story Media, Inc. v. OpenAI, Inc.*, 756 F. Supp. 3d 1 (S.D.N.Y. 2024).

228. *N.Y. Times Co. v. Microsoft, Corp.*, 777 F. Supp. 3d 283 (S.D.N.Y. 2025).

229. *Tremblay v. OpenAI, Inc.*, 716 F. Supp. 3d 772 (N.D. Cal. 2024).

230. Kevin Madigan, *AI Lawsuit Developments in 2024: A Year in Review*, COPYRIGHT ALL. (Jan. 9, 2025), <https://perma.cc/R4Z2-L3JM> (“The proliferation of generative artificial intelligence (GAI) models over the past few years has given rise to well over thirty copyright infringement lawsuits by copyright owners against GAI developers.”).

231. See *supra* note 29 (describing the inability of LLM creators to keep CMI paired with data throughout the training process); see also Kahveci, *supra* note 30, at 798-99 (2023) (“Giving attribution to works that are used to create a certain output requires additional extensive work by the AI systems as it is not easy to determine which specific works are used for a specific output.”); Henderson et al., *supra* note 23, at 19 (“[F]oundation models rarely, if ever, provide proper attribution Indeed, in many cases, it can be difficult to determine which training examples actually contributed to a given generation.”).

available defenses, and the statutory damages scheme—work together to make it exceptionally easy for plaintiffs to state § 1202 claims.²³²

Certainly, these problems have not yet resulted in a deluge of judgments against LLM developers. Of the seven relevant cases brought against LLMs,²³³ all involved the dismissal of at least one § 1202 claim. But absent doctrinal changes, there is little reason to expect this trend to continue. First, the precedent here is scarce. The theoretical ease of making out a claim may not play itself out in every case, and all seven cases are concentrated in only two district courts. Second, the dismissals do not show a systematic difficulty in alleging a § 1202 violation.²³⁴ There was no discernable pattern across cases as to the reason for dismissal, suggesting that no one requirement of a § 1202 claim is particularly difficult to overcome. Both *Doe v. GitHub* and *Andersen v. Stability* dismissed the plaintiffs' claims for lack of identity,²³⁵ but most courts outside the Ninth Circuit do not subscribe to the identity requirement.²³⁶ While the *New York Times v. Microsoft* court dismissed § 1202(b)(1) claims against OpenAI due to insufficient evidence of intentional

232. On October 16, 2025, I ran the following searches on Westlaw: (1) ["copyright infringement"], filtering the jurisdiction to all federal cases and the cause of action to "copyright violation: general infringement" and limiting the date range to cases before April 5, 2025, and (2) ["digital millennium copyright act" OR "DMCA") AND "1202"], filtering the jurisdiction to all federal cases and the cause of action to "copyright violation: digital millennium copyright violation" and limiting the date range to cases before April 5, 2025 (the latest date available as the upper bound of the date range). The first search revealed that of 419 copyright infringement cases in which a motion to dismiss for failure to state a claim was brought, there were 362 motions granted and 192 denied (with some cases including multiple claims). The second search revealed that of the 36 § 1202 DMCA cases in which a motion to dismiss for failure to state a claim was brought, 29 were denied and 27 were granted (again, with some cases including multiple claims). This comparison is a rough demonstration of how much more likely copyright infringement claims are to be dismissed than DMCA § 1202 claims: Whereas infringement claims were dismissed at nearly twice the rate at which they survived (approximately a 1.9-to-1 dismissal-to-survival ratio), § 1202(b) claims survived at markedly higher rates (a near 1-to-1 ratio). Indeed, among the cases reviewed, motions to dismiss § 1202(b) claims were slightly more likely to be denied than granted.

233. On October 16, 2025, I ran the following search on Westlaw: ["digital millennium copyright act" OR "DMCA") AND "1202" AND ("large language model" OR "artificial intelligence")]. Seven of the results represented relevant and distinct cases. See cases cited *infra* notes 227, 235, 237-42.

234. See Crusey, *supra* note 18, at 507 ("The fact-specific nature of Section 1202(b) claims makes it difficult (and perhaps unwise) to infer from these suits how other Section 1202(b) claims will progress.").

235. *Doe v. GitHub, Inc.*, No. 22-cv-06823, 2024 WL 235217, at *8 (N.D. Cal. Jan. 22, 2024); *Andersen v. Stability AI Ltd.*, 744 F. Supp. 3d 956, 971 (N.D. Cal. 2024).

236. See *supra* note 104 and accompanying text (explaining which courts implement an identity requirement).

removal of CMI,²³⁷ the court in *Intercept v. OpenAI* found the plaintiff in that case plausibly alleged the same.²³⁸ And while the court in *Tremblay v. OpenAI* dismissed the plaintiff's § 1202(b)(1) claims in part because the second scienter requirement was insufficiently pleaded,²³⁹ the court in *Concord Music Group v. Anthropic* found plausible allegations of the requisite scienter.²⁴⁰

Dicta from these cases also suggest that future matters may be decided differently. The court in *Andersen v. Stability* clarified that the plaintiffs resolved the court's previous reasons for dismissal, suggesting that absent the identity requirement, Andersen's claims might not have been dismissed.²⁴¹ The court in *GitHub* suggested the same.²⁴²

Should future plaintiffs note these decisions and the current identity-similarity split, they can take advantage of the low bar created by § 1202's elements in the many courts that do not impose an identity requirement.

Weaker Intent.—As discussed in Part II.D, the first scienter requirement states that the violative action itself must be intentional (for § 1202(b)(1)) or knowing (for § 1202(b)(2) and (3)), and the second scienter requirement requires “actual or constructive knowledge” that a violative act “will induce, enable, facilitate, or conceal an infringement.”²⁴³ Multiple courts have treated the act of removing CMI as sufficient evidence of intent to remove CMI.²⁴⁴ This pattern has continued as applied to LLMs, with one court finding that an LLM provider's use of CMI removal tools was enough to allege intent under the first prong of § 1202.²⁴⁵

In *New York Times v. Microsoft*, the plaintiffs alleged, among other claims, that OpenAI violated § 1202(b)(1) by removing CMI from its training datasets, which plaintiffs alleged included their works.²⁴⁶ Two groups of plaintiffs—the

237. *N.Y. Times Co. v. Microsoft Corp.*, 777 F. Supp. 3d 283, 314 (S.D.N.Y. 2025).

238. *Intercept Media, Inc. v. OpenAI, Inc.*, 767 F. Supp. 3d 18, 29 (S.D.N.Y. 2025).

239. *Tremblay v. OpenAI, Inc.*, 716 F. Supp. 3d 772, 779 (N.D. Cal. 2024).

240. *Concord Music Grp., Inc. v. Anthropic PBC*, No. 24-cv-03811, slip op. at 10 (N.D. Cal. filed Apr. 25, 2025), <https://perma.cc/27CV-ZEM4> (denying a motion to dismiss the plaintiffs' amended complaint).

241. *See Andersen v. Stability AI Ltd.*, 744 F. Supp. 3d 956, 970 (N.D. Cal. 2024) (confirming that “as directed in the [previous dismissal],” the plaintiffs “identified the CMI present on their works that they contend has been stripped by Stability” and claimed that “Stability engaged in knowing removal of CMI . . . based on plausible allegations”).

242. *See Doe v. GitHub, Inc.*, 672 F. Supp. 3d 837, 858 (N.D. Cal. 2023) (finding, before dismissing the complaint for lack of identity, that the plaintiffs pleaded “sufficient facts to support a reasonable inference that Defendants intentionally designed the programs to remove CMI from any licensed code they reproduce as output”).

243. *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 171 (2d Cir. 2020) (quoting 17 U.S.C. § 1202(b)(3)).

244. *Supra* notes 201-04 and accompanying text.

245. *N.Y. Times Co. v. Microsoft Corp.*, 777 F. Supp. 3d 283, 314-15 (S.D.N.Y. 2025).

246. *Id.* at 309-10.

Daily News and the Center for Investigative Reporting—alleged in their complaints that when OpenAI scraped their content off the internet to create datasets used to train its LLMs, it “intentionally used the Dagnet and Newspaper content extractors, which are designed to remove CMI from the works they scrape from the internet.”²⁴⁷ OpenAI’s use of these tools to extract content from the web, in combination with its employment of “highly skilled data scientists who would know how Dagnet and Newspaper work,” was enough to plausibly allege intentional removal of CMI.²⁴⁸ Again, showing removal of CMI was enough to show intention to remove CMI.

On the other hand, in the same case, the New York Times’ complaint “fail[ed] to plausibly allege that OpenAI removed CMI from the training datasets” at all.²⁴⁹ Its lack of “specific detail on how CMI was allegedly removed”²⁵⁰ led the court to grant OpenAI’s motion to dismiss the New York Times’ claim. And of course, not all courts have held that removal of CMI is enough to satisfy the first scienter prong: Where Anthropic was accused of training an LLM on a plaintiff’s work and removing the plaintiff’s CMI, the Northern District of California explicitly found that “simply [repeating] the statutory language without stating any facts to show that Anthropic intentionally removed CMI” was insufficient to satisfy the first scienter prong.²⁵¹

This inconsistency in courts’ outcomes should hardly reassure LLM providers, though: The sufficiency of these allegations in *any* court where there is such little precedent applying § 1202 to defendant LLM providers both affects the risk calculus of companies creating LLMs and forecasts future difficulty for LLM providers at the motion to dismiss stage.

The second scienter requirement spells further trouble for LLM providers as the infringement they are alleged to induce, enable, facilitate, or conceal may very well be their own. As discussed in Part II.D, a defendant’s actual or

247. *Id.* at 315. Note that Dagnet is a tool that uses machine learning algorithms that “extract the main article content” from web pages. See *Dagnet*, GITHUB, <https://perma.cc/6BVD-K4KA> (archived Dec. 31, 2025). Newspaper is a similar tool that “specifically target[s] news article pages.” See Janek Bevendorff, Sanket Gupta, Johannes Kiesel & Benno Stein, *An Empirical Comparison of Web Content Extraction Algorithms*, 2023 PROCS. 46TH INT’L ACM SIGIR CONF. ON RSCH. & DEV. INFO. RETRIEVAL 2594, 2598, <https://perma.cc/JM4W-PEQ2>.

248. *N.Y. Times*, 777 F. Supp. 3d at 315 (quoting First Amended Complaint ¶ 64, *Ctr. for Investigative Reporting, Inc. v. OpenAI, Inc.*, No. 24-cv-04872 (S.D.N.Y. Apr. 4, 2025) (Nos. 23-cv-11195, 24-cv-03285 & 24-cv-04872), ECF No. 88).

249. *Id.* at 314.

250. *Id.*

251. *Concord Music Grp., Inc. v. Anthropic PBC*, No. 24-cv-03811, 2025 WL 1487988, at *6 (N.D. Cal. Mar. 26, 2025). The court later found sufficient allegations of scienter based on an amended complaint. See *supra* note 240.

constructive knowledge of their own infringement is enough to satisfy the second scienter requirement.²⁵² Because LLMs' outputs are inherently distributed to a user, if an infringement defense does not apply, the simple act of producing the output at issue may be enough—and has been shown to be enough—to satisfy this requirement.²⁵³

Weaker Standard for Proving Copying.—As discussed in Part II.A, the standard for proving a work is a copy under § 1202 is permissive compared to the standard for infringement claims. The standard under infringement requires both access and similarity or striking similarity to prove copying occurred *and* considers copyrightability in determining whether that copying is legally sufficient.²⁵⁴ The bar is much lower for some courts hearing § 1202 claims.²⁵⁵ In particular, it is extremely difficult for LLM providers to establish compliance in courts that adhere to a similarity standard.²⁵⁶

Current large-scale generative AI training practices routinely entail the separation of works from their attendant copyright management information.²⁵⁷ This separation is a de facto feature of most modern training

252. *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 172 (2d Cir. 2020) (“[A]n infringement’ is not limited to the infringing acts of third parties. The plain meaning of the statutory language also encompasses an infringement committed by the defendant himself. This includes the knowing, unauthorized infringement that serves as the basis of establishing the first scienter element of Section 1202(b).”).

253. *N.Y. Times*, 2025 WL 1009179, at *17 (“[P]laintiffs have plausibly alleged both that OpenAI’s removal of CMI conceals its *own* infringement, and that that removal enables and facilitates third-party infringement.”); *Intercept Media, Inc. v. OpenAI, Inc.*, 767 F. Supp. 3d 18, 31 (S.D.N.Y. Feb. 20, 2025) (finding that ChatGPT’s ability to generate regurgitations is a persuasive theory of downstream infringement).

254. See *supra* notes 77-89 and accompanying text (describing the standard for proving copying for a copyright infringement claim).

255. See *supra* notes 122-44 and accompanying text (describing the similarity standard and § 1202’s lack of a copyrightability inquiry).

256. Cases establishing a similarity standard tend to hold in favor of plaintiffs. See *Oracle Int’l Corp. v. Rimini St., Inc.*, No. 14-cv-01699, 2023 WL 4706127, at *82 (D. Nev. July 24, 2023) (rejecting the identity requirement and finding for plaintiff); *ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d 411, 424, 427-28 (S.D. Tex. 2023) (rejecting the identity requirement and finding plaintiff stated a claim for a § 1202 violation); *New Parent World, LLC v. True to Life Prods., Inc.*, No. CV-23-08089, 2024 WL 4277865, at *2 (D. Ariz. Sept. 24, 2024) (rejecting the identity requirement and denying defendant’s motion to dismiss); *Real World Media LLC v. Daily Caller, Inc.*, 744 F. Supp. 3d 24, 40-41 (D.D.C. 2024) (same); *Enter. Tech. Holdings v. Noveon Sys., Inc.*, No. 05-CV-2236, 2008 WL 11338356, at *7 (S.D. Cal. July 29, 2008) (employing substantial similarity and granting the plaintiff’s motion for partial summary judgment).

257. Boyden, *supra* note 29, at 278-79 (“Obviously the training process for a generative AI model is not intended to preserve a work in its entirety, including any CMI contained in the work itself or a file’s metadata. . . . But training an AI model by its nature scrambles the plaintiff’s works into a disparate set of associated data points.”); Samuelson et al., *supra* note 29, at 8 (“The generative AI training process extracts
footnote continued on next page”)

pipelines.²⁵⁸ If the output generated without this CMI resembles copyrighted training data, LLM providers may be liable for removing CMI under § 1202(b)(1) and for distributing work without CMI under § 1202(b)(3).²⁵⁹

Because the training process separates works into parts and aggregates them before extracting information,²⁶⁰ the only option for an output that resembles a copyrighted work to comply with § 1202 would be to match the output with the training data that contributed to it, along with that data's associated CMI.²⁶¹ This process of "[c]onnecting the AI-generated content to both the individuals responsible for its generation and to the training data that allows for its generation is called attribution,"²⁶² and it is difficult for LLMs to achieve.²⁶³

One reason for this difficulty is that multiple pieces of training data may be similar or identical to one another, making it difficult to determine exactly which data contributed to the output.²⁶⁴ This issue is especially salient in the

information from millions or billions of works and, in the process, disassembles or tokenizes their elements to construct a very different representation in the models.").

258. *Id.*; see *supra* note 29 (describing the difficulty of associating CMI to training data). But note that it may nonetheless be *possible* to train an LLM on data without separating that data from its CMI. For this reason, courts do not automatically assume that CMI was removed during the training process. See, e.g., *N.Y. Times Co. v. Microsoft Corp.*, 777 F. Supp. 3d 283, 314 (S.D.N.Y. 2025) ("The Times contends that because the regurgitating outputs listed in their complaints lack CMI, then *a fortiori* CMI was removed by defendants during the training process. . . . [I]t is entirely plausible that CMI remained on the articles included in the training datasets but simply did not appear in the outputs.").

259. See Second Amended Complaint, *supra* note 12, at 17 ("[T]he Output is often a near-identical reproduction of code from the training data."); *id.* at 53-54 (claiming violations of § 1202(b)(1) and § 1202(b)(3)); *Doe v. GitHub, Inc.*, 672 F. Supp. 3d 837, 857-58 (N.D. Cal. 2023) (finding plaintiffs sufficiently pleaded intentionality by alleging that the defendants knew CMI was attached to the training data and trained the model to stop reproducing it).

260. Boyden, *supra* note 29, at 272-73; Samuelson et al., *supra* note 29, at 8.

261. See 26 U.S.C. § 1202(b)(1), (3).

262. Emanuele Mezzi et al., *Who Owns the Output? Bridging Law and Technology in LLMs Attribution*, ARXIV 2504.01032, at 1 (2025) (emphasis omitted), <https://perma.cc/W3EQ-NLG5>.

263. Jingtang Wang, Xinyang Lu, Zitong Zhao, Zhongxiang Dai, See-Kiong Ng & Bryan Kian Hsiang Low, *Source Attribution for Large Language Model-Generated Data*, ARXIV 2310.00646v2, at 2 (2024), <https://perma.cc/3UR2-G8BL> ("[E]ffective source attribution for LLMs remains an open problem." (emphasis omitted)).

264. Kahveci, *supra* note 30, at 798-99 ("Giving attribution to works that are used to create a certain output requires additional extensive work by the AI systems as it is not easy to determine which specific works are used for a specific output. For example . . . the training dataset of ChatGPT may contain three different documents that have the relevant paper [S]cientific studies show that it is challenging for the generative AI to determine which one of these three documents was the source of that output.").

context of *GitHub*, where the LLM at issue, Copilot, is a code completion tool.²⁶⁵ Code is a constrained form of expression, and it is thus likely that a snippet of code expressing a given idea will contain similarities with other programs expressing that same idea.²⁶⁶ In other words, there are only so many ways to code certain functions. Code, like much of the data LLMs are trained on, is functional,²⁶⁷ and thus it may be saved from copyright infringement by the merger doctrine (which withholds protection where there are few ways of saying something) or the idea-expression dichotomy (which withholds protection where the thing being said is an idea rather than an expression—like a fact).²⁶⁸ In any event, the result is a non-unique piece of data whose similarity to a given output is not necessarily indicative of copying.²⁶⁹

Like code, much of the data LLMs are trained on is thinly copyrighted.²⁷⁰ However, because the § 1202 analysis does not consider the copyrightability of a work in determining the proper standard for liability, works that are subject to either thin or thick copyright protection are subject to the similarity standard in the courts in which that standard applies.²⁷¹ For thinly copyrighted works, this standard is improper; traditional copyright doctrine provides that the proper standard for proving copying for thinly copyrighted

265. *Inline Suggestions from Github Copilot in VS Code*, VISUAL STUDIO CODE, <https://perma.cc/H8KG-ZLBW> (archived Dec. 31, 2025) (describing Copilot’s code completion functionality).

266. Stephen R. Mick, *Applying the Merger Doctrine to the Copyright of Computer Software*, 37 COPYRIGHT L. SYMP. 173, 206 (1987) (“[T]he structure and organization of two programs that are designed to accomplish the same goal will almost certainly and sometimes necessarily be similar.”).

267. Charles Duan, *What Is Copyrightable in Software?* 94 GEO. WASH. L. REV. (forthcoming 2026) (manuscript at 14-15, 15 n.68), <https://perma.cc/3RFV-W6R5>.

268. See *supra* Part II.A (explaining various copyrightable subject matter doctrines).

269. 4 NIMMER & NIMMER, *supra* note 22, § 13D.37(D)(b)(v)(I) (“The mere fact that similar practices or techniques are employed in two programs is tenuous evidence of copying, at best.”).

270. Consider the makeup of the Common Crawl, for example, a popular database for LLM training. Stefan Baack, *A Critical Analysis of the Largest Source for Generative AI Training Data: Common Crawl*, 2024 PROCS. ACM CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 2199, 2199, <https://perma.cc/GDY2-RTPA> (“A majority of LLMs published both before and since [GPT-3] by other developers likewise rely heavily on filtered versions of Common Crawl for their pre-training.”). As of December 2025, almost 20% of the Common Crawl is from Wikipedia, and therefore likely predominantly factual content. *Top-500 Registered Domains of the Latest Main Crawl*, COMMON CRAWL, <https://perma.cc/TQW2-75JP> (archived Dec. 31, 2025). Wiktionary contributes an additional 4.8%. *Id.* Across the Common Crawl, the number of news articles, scientific literature, product manuals, legal texts, and forum discussions—while challenging to estimate—are likely substantial as well.

271. See *supra* Part II.A (explaining § 1202’s lack of a copyrightability analysis).

works is identity.²⁷² A less stringent standard risks finding liability for works that were independently generated, or that are not sufficiently protectable to warrant liability.

Though this issue is especially salient where the subject matter of the output is thinly copyrighted, like code or a factual compilation, it applies to any attempt at attribution. Even where the merger doctrine and the idea-expression dichotomy are not applicable to a particular output—that is, even where an output is not thinly copyrighted—similarity to training data is possible by sheer coincidence. This is especially true given the large corpus of works on which LLMs train.

In addition, courts that follow the similarity standard make attributing LLM output uniquely difficult because it is far easier for LLMs to check and filter for identical outputs than it is for them to prevent similar outputs.²⁷³ Technology exists to identify regurgitated training data. Indeed, this technology has already been implemented: Copilot offers users a tool that “display[s] references to” code suggestions of a certain length that match public code,²⁷⁴ and Google announced its Bard tool would cite webpages where it “directly [quotes the webpage] at length.”²⁷⁵ Similarity, however, is harder to identify. This is in part due to the lack of specificity and consistency with which courts have applied the similarity standard. But it is also a technological constraint: Even methods deemed similarity checks function by identifying matching content²⁷⁶—that is, they check for identity.²⁷⁷

272. 4 NIMMER & NIMMER, *supra* note 22, § 13D.32(A) (adding that multiple circuits agree with this approach and that “virtual identity is not the same as absolute identity”).

273. Henderson et al., *supra* note 23, at 26 (“[M]ost technical work on copyright evaluation and mitigation focuses on near-verbatim overlap”); see *infra* notes 274-77 and accompanying text (describing current technology for preventing identical outputs and identifying technology that is ostensibly preventing similarity as in fact preventing identity).

274. *Finding Public Code That Matches GitHub Copilot Suggestions*, GITHUB DOCS, <https://perma.cc/P39L-YWN7> (archived Dec. 31, 2025) (referring to users’ ability to toggle “suggestions that match publicly available code”); *GitHub Copilot Code Referencing*, GITHUB DOCS, <https://perma.cc/55ZG-QHX5> (archived Dec. 31, 2025) (specifying that 150 characters of code are compared).

275. Kahveci, *supra* note 30, at 799 n.20 (“If Bard does directly quote at length from a webpage, it cites that page.” (quoting *Bard FAQ*, BARD EXPERIMENT, <https://perma.cc/G2V6-4SNS> (archived Apr. 19, 2023))).

276. See, e.g., Patti West-Smith, *Does Turnitin Detect Plagiarism?*, TURNITIN (Oct. 5, 2022), <https://perma.cc/9X9D-UT4G>.

277. Kahveci, *supra* note 30, at 799 (“Indeed, this kind of attribution can be done through a basic similarity check, which is already employed by plagiarism detection products like Turnitin.”); *Understanding the Similarity Score*, TURNITIN, <https://perma.cc/55ZV-QKQ9> (archived Feb. 2, 2026) (“The similarity score is simply the percentage of text in a submission that matches other sources.”).

Still No Fair Use.—As the biggest hurdle standing in the way of widespread infringement liability for LLM developers, fair use “may be the key to determining how AI will develop.”²⁷⁸ And it is not only the potential monetary damages that might be prohibitive to the development of generative AI: “Some warn that requiring AI companies to license copyrighted works would throttle a transformative technology, because it is not practically possible to obtain licenses for the volume and diversity of content necessary to power cutting-edge systems.”²⁷⁹ In short, fair use enables generative AI.

A multitude of cases are either currently pending or recently decided regarding copyright infringement of AI models; many defendants are expected to bring—or have already brought—a fair use defense.²⁸⁰ The centrality of fair use to the legality of generative AI makes its potential inapplicability to § 1202 powerful.²⁸¹

278. Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 FORDHAM L. REV. 1887, 1895 (2024).

279. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE PART 3: GENERATIVE AI TRAINING 1 (Pre-Publication Version, May 2025).

280. See, e.g., *Andersen v. Stability AI Ltd.*, 744 F. Supp. 3d 956, 983 (N.D. Cal. 2024) (“DeviantArt contends that any use of plaintiffs’ works in Stable Diffusion 1.4 should be considered fair use.”); Defendant Anthropic PBC’s Notice of Motion and Motion for Summary Judgment at viii, *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007 (N.D. Cal. 2025) (3:24-CV-05417), 2025 WL 1304762, ECF No. 122 (“The Motion should be granted, and summary judgment entered in Anthropic’s favor because Anthropic’s use of plaintiffs’ works is a fair use under Section 107 of the Copyright Act.”); *Concord Music Grp., Inc. v. Anthropic PBC*, 738 F. Supp. 3d 973, 979 (M.D. Tenn. 2024) (“Anthropic maintains that its inclusion of copyrighted song lyrics in the training corpus for Claude constitutes ‘fair use’ under 17 U.S.C. § 107, and, therefore, Anthropic is not required to pay for licenses over those lyrics.”); *Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc.*, 765 F. Supp. 3d 382, 401 (D. Del. 2025) (“I grant summary judgment for Thomson Reuters on fair use.”); Answer at 7, *Dubus v. NVIDIA Corp.*, No. 24-cv-02655 (N.D. Cal. July 1, 2024) (Nos. 24-cv-01454 & 24-cv-02655), ECF No. 48 (asserting fair use as a third affirmative defense); *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1060 (N.D. Cal. 2025) (“Therefore, on this record, Meta is entitled to summary judgment on its fair use defense to the claim that copying these plaintiffs’ books for use as LLM training data was infringement.”); *In re Google Generative AI Copyright Litig.*, No. 23-cv-03440, 2025 WL 1159998, at *3 (N.D. Cal. Apr. 21, 2025) (“[T]he Court cannot determine who is a member of the class without deciding the merits of each potential class member’s claim, including . . . whether Google has a valid defense based on fair use or license (among other possible defenses.)”); Complaint ¶ 75, *Dow Jones & Co. v. Perplexity AI, Inc.*, 797 F. Supp. 3d 305 (S.D.N.Y. 2025) (No. 24 Civ. 7984), 2024 WL 4564378, ECF No. 1 (“At other times, Perplexity has publicly alluded to the possibility that its copying of publishers’ copyrighted works could possibly be justified as a fair use.”); see also Ryan Browne, *Getty Images Spending Millions to Battle a ‘World of Rhetoric’ in AI Suit, CEO Says*, CNBC (updated May 28, 2025, 6:55 AM EDT), <https://perma.cc/4YV3-V64B> (“[Stability AI] has previously argued its use of copyright-protected material online is sound under the ‘fair use’ doctrine . . .”).

281. See *supra* Part II.B.

Section 1202 claims have been relatively underutilized compared to infringement claims. In 2025, federal courts cited § 1202 in about 900 opinions,²⁸² while infringement has been cited in over 6,000.²⁸³ And when § 1202 claims do appear, they rarely stand alone; instead, they have often been brought alongside infringement. In fact, of the forty-five opinions published as of this writing, only two include § 1202 claims brought alone.²⁸⁴ In all but one of the remaining cases, § 1202 claims were accompanied by a copyright infringement claim.²⁸⁵ This suggests that § 1202 is currently viewed by plaintiffs as a supplemental claim to infringement.

How and whether courts choose to apply fair use to copyright infringement claims against LLM providers could change the way § 1202 claims are brought. If courts decide that fair use applies to the acts of LLM providers—as one court has recently held²⁸⁶—§ 1202 claims could become attractive to plaintiffs as a workaround. That is, if fair use continues to apply as a hurdle to plaintiffs’ infringement claims, and if courts continue to decline fair use’s application to § 1202 claims,²⁸⁷ plaintiffs may avoid fair use completely by alleging a § 1202 violation—not in addition to an infringement claim, but in its stead.

One court’s recent decision seems to make this outcome less likely. In *Kadrey v. Meta*, the Northern District of California not only decided that fair use applied to the plaintiffs’ copyright infringement claim against Meta for Meta’s training of its LLM on copyrighted data²⁸⁸—the court also decided that

282. Citing References for 17 U.S.C. § 1202, 921 results (Dec. 31, 2025), WESTLAW (on file with the *Stanford Law Review*) (filtered by “Cases”).

283. Citing References for 17 U.S.C. § 501, 6,197 results (Dec. 31, 2025), WESTLAW (on file with the *Stanford Law Review*) (filtered by “Cases”).

284. *See supra* note 280 (referencing the § 1202 opinions published in 2025). For cases that brought a § 1202 claim alone, see, for example, *Webster v. Doe*, No. 25-CV-931, 2025 WL 2054350, at *1 (E.D.N.C. July 22, 2025); and *Intercept Media, Inc. v. OpenAI, Inc.*, 767 F. Supp. 3d 18, 22 (S.D.N.Y. 2025).

285. *See supra* note 282 (referencing the § 1202 opinions published in 2025); *Crafts Outlet, Inc. v. Crafts Online Outlet, Inc.*, No. 23-cv-02278, 2025 WL 2197314, at *3 (D. Colo. Aug. 1, 2025) (involving a § 1202 claim brought alongside a trademark infringement claim but not a copyright infringement claim).

286. *See Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1034 (N.D. Cal. 2025) (“This order grants summary judgment for Anthropic that the training use was a fair use.”); *see also Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1060 (N.D. Cal. 2025) (granting the defendant’s motion for summary judgment on its fair use defense). Note, however, that the court emphasized the fact-specific nature of the fair use inquiry in *Kadrey* and noted in dicta that “it’s hard to imagine that it can be fair use to use copyrighted books to develop” LLM tools. *Id.* at 1059.

287. *See supra* Part II.B (discussing the potential inapplicability of fair use to § 1202 claims).

288. *Kadrey*, 788 F. Supp. 3d at 1036-37.

fair use precluded the plaintiffs' accompanying § 1202 claim.²⁸⁹ Though this decision seemingly subjects § 1202 claims to the fair use hurdle, it is in reality an extremely fact-specific outcome that does not foretell a general application of fair use to § 1202.

In *Kadrey*, the § 1202 claim that survived the motion to dismiss stage was a very narrow one: The plaintiffs alleged that Meta “knew or should have known that removing CMI from the plaintiffs’ books would conceal infringement by making [its LLM] less likely to regurgitate CMI and thus indicate that it was trained on copyrighted material.”²⁹⁰ Because the court held that training on copyrighted material constituted fair use, there was no underlying infringement to conceal. Fair use thus properly foreclosed the § 1202 claim in this instance. That outcome does not mean, however, that fair use is generally an independent bar to § 1202 claims; rather, it reflects that here, plaintiffs’ allegation was so specific that the application of fair use meant that the alleged infringement was not actually an infringement. Had plaintiffs alleged instead that Meta’s removal of CMI facilitated third-party infringement, fair use’s application to the copyright infringement claim would not have precluded the § 1202 claim. While *Kadrey* illustrates the rare circumstance in which fair use can incidentally foreclose a § 1202 claim, in most cases fair use provides no defense.²⁹¹

Even More Damages.—Compared to statutory damages for infringement, which accumulate “per work,” § 1202’s “per-violation” scheme may disadvantage LLM providers. This is because such defendants may be accused of violating various provisions of § 1202 (i.e., § 1202(b)(1) for removing CMI and § 1202(b)(3) for distributing outputs that resemble training data sans CMI), which could double the number of damage awards, and because LLMs are uniquely situated for widespread use. While distribution of a single violative act to multiple people does not qualify as multiple violations, LLMs create distinct, independent outputs for each user.²⁹² This means violations stack up

289. *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417, 2025 WL 1786418, at *1 (N.D. Cal. June 27, 2025) (“The plaintiffs’ DMCA claim fails because . . . [defendant’s] copying must be deemed fair use . . .”).

290. *Id.* at *1 n.1.

291. See *supra* Part II.B (discussing the potential inapplicability of fair use to § 1202 claims).

292. See Chesley, *supra* note 26, at 34 (“In *McClatchey*, the court held that each ‘violative act [performed] by the Defendant’ constituted a violation. This is an arbitrary construct when there is little difference between distributing a photograph to a list of subscribers via one photo database or a series of several emails. However, under the per-violation scheme of the DMCA, we have little choice but to use such language.” (quoting *McClatchey v. Associated Press*, No. 5-cv-145, 2007 WL 1630261, at *6 (W.D. Pa. June 4, 2007))); Shuyin Ouyang, Jie M. Zhang, Mark Harman & Meng Wang, *An Empirical Study of the Non-Determinism of ChatGPT in Code Generation*, 34 ACM TRANSACTIONS ON
footnote continued on next page

quickly. For example, if one in every 10,000 generative AI model outputs is a regurgitation of copyrighted material, one million yearly users each generating only ten outputs a year comes out to \$25 billion in damages. And these numbers are extremely conservative for defendants like OpenAI, with ChatGPT now approaching 700 million users *per week*.²⁹³

These staggering damages amounts are not just theoretical. The plaintiffs in *GitHub* explained the \$9 billion in § 1202 statutory damages they sought as comprising the \$2,500 statutory minimum multiplied by three for each of the (at the time) 1.2 million Copilot users.²⁹⁴ They explained that “[e]ach time Copilot provides an unlawful Output it violates [the CMI rules] three times (distributing the Licensed Materials without: attribution, copyright notice, and License Terms).”²⁹⁵ Assuming that each user generated only a single output violating § 1202, the number of violations would be 3,600,000.²⁹⁶ When multiplied by \$2,500, the plaintiffs landed at \$9 billion in damages.²⁹⁷

A per-work damages scheme, on the other hand, would limit the damages multiplier. For a § 1202 violation, the number of works being violated is limited: The output must meet the statutory definition of a “copy.”²⁹⁸ Regardless of whether the standard used is similarity or identity, a § 1202 plaintiff must, as a matter of common sense, allege that the output is attributed primarily to a particular work—i.e., that the output is a verbatim or sufficiently similar “copy” of the plaintiff’s work.²⁹⁹ The necessity of arguing that the output is a “copy” of the plaintiff’s work to state a claim inherently limits the damages multiplier: Though courts interpret the specifics of each standard differently, as a general rule, the more works to which an output owes attribution, the more likely it is to fail to satisfy a similarity or identity standard. Thus, even 1,000 violating LLM outputs would likely be limited to a single-digit multiplier under a per-work scheme, since the output, by definition, must draw primarily from the plaintiff’s work. But under a per-violation scheme, the same 1,000 violating outputs could be subject to a

SOFTWARE ENG’G & METHODOLOGY, art. 42, at 1 (2025) (“[LLMs] are non-deterministic by nature.”).

293. MacKenzie Sigalos, *OpenAI’s ChatGPT to Hit 700 Million Weekly Users, Up 4x from Last Year*, CNBC (updated Aug. 4, 2025, 2:46 PM EDT), <https://perma.cc/A4SV-8U8H>.

294. Pamela Samuelson, Opinion, *How to Think About Remedies in the Generative AI Copyright Cases*, COMMS. ACM, July 2024, at 27, 28.

295. *Id.*

296. *Id.*

297. *Id.*

298. See *supra* notes 97-98 and accompanying text.

299. See *supra* Part IIA (describing how courts have understood the similarity and identity standards).

practically infinite multiplier, depending on how many users generated that output and how many times each user did so.

In addition, the difficulty of assessing damages is a factor courts consider in choosing high damages awards,³⁰⁰ and this factor may work to the detriment of defendant LLM providers. LLM providers have been the subject of much criticism regarding the secrecy of their models and training processes.³⁰¹ This secrecy has been justified on commercial grounds,³⁰² but it contributes to uncertainty regarding how many alleged violations may be embedded into the training and output generation process.³⁰³ Information about how many users

300. See *supra* note 193.

301. See, e.g., First Amended Complaint at 17-18, *N.Y. Times Co. v. Microsoft Corp.*, 777 F. Supp. 3d 283 (S.D.N.Y. 2025) (Nos. 23-cv-11195) (“GPT-3.5 and GPT-4 are both orders of magnitude more powerful than the two previous generations, yet Defendants have kept their design and training entirely a secret. . . . OpenAI’s Chief Scientist Sutskever justified this secrecy on commercial grounds But its effect was to conceal the identity of the data OpenAI copied to train its latest models from rightsholders”); First Amended Complaint, *supra* note 111, ¶ 9 (“Though Defendants have been cagey about what data was used to train the AI, they have conceded that the training data includes data in vast numbers of publicly accessible repositories on GitHub” (footnotes omitted)); Andreas Liesenfeld, Alianda Lopez & Mark Dingemans, *Opening Up ChatGPT: Tracking Openness, Transparency, and Accountability in Instruction-Tuned Text Generators*, 2023 PROCS. 5TH INT’L CONF. ON CONVERSATIONAL USER INTERFACES, art. 47, § 1.1, <https://perma.cc/9A5P-9A28> (criticizing models like ChatGPT for preventing a “critical and holistic understanding of the constraints and capabilities of such systems”); Camilla A. Hrdy, *Keeping ChatGPT a Trade Secret While Selling It Too*, 40 BERKELEY TECH. L.J. 75, 79-80 (2025) (“[Many AI companies] distribute generative AI models using a ‘closed source’ structure that hides the model’s inner workings from users. ChatGPT users, for example, generally have no idea how the underlying generative AI models work or how they reach their conclusions. . . . From the users’ perspective, ChatGPT is the ultimate ‘black box.’” (footnote omitted) (quoting Saurabh Bagchi, *What Is an AI ‘Black Box?’*, GIZMODO (May 28, 2023), <https://perma.cc/WF17-Q642>; *ChatGPT Is a Black Box: How AI Research Can Break It Open*, 619 NATURE 671, 671-72 (2023); and Daniel Hardt, *Everyone Uses ChatGPT, but It Is a Black Box: Here’s How to Make the Most of It*, COPENHAGEN BUS. SCH. (Oct 10, 2023), <https://perma.cc/9D8L-99P8>)).

302. Adam Buick, *Copyright and AI Training Data—Transparency to the Rescue?*, 20 J. INTELL. PROP. L. & PRAC. 182, 184 (2025) (“For its part, OpenAI justified its decision not to release further details regarding GPT-4 on the basis of concerns regarding ‘the competitive landscape and the safety implications of large-scale models’ OpenAI’s Chief Scientist . . . clarified that OpenAI believes that sharing further details regarding their training data would facilitate the replication of their cutting-edge AI models, while releasing detailed information regarding training data would enable careless or malicious actors to develop their own powerful AI models more easily.” (quoting OpenAI, *GPT-4 Technical Report*, ARXIV 303.08774v6, at 2 (2024), <https://perma.cc/8M5C-RUET>)).

303. Some plaintiffs have accused LLMs of violating § 1202 at the output stage. See *Doe v. GitHub, Inc.*, 672 F. Supp. 3d 837, 858 (N.D. Cal. 2023) (“Plaintiffs thus plead sufficient facts to support a reasonable inference that Defendants intentionally designed the programs to remove CMI from any licensed code they reproduce as output.”). Other
footnote continued on next page

have generated a given output is also obscured; unlike a traditional § 1202 claim where, for example, the plaintiff might be able to ascertain the number of purchasers of a violating book or the number of digital downloads of a violating song, there is no publicly available data on how many times a certain output has been generated by an LLM.³⁰⁴ Courts that encounter these difficulties in assessing damages may levy high damages awards.

IV. Bridging Purpose and Practice

Implementing an identity requirement for LLMs would mitigate the overbreadth of § 1202 without the underinclusiveness of piecemeal reforms. Such a carveout is supported by precedent—§ 1202 currently exempts broadcasters where it is not “technically feasible” to avoid a § 1202(b) violation or where compliance would “create an undue financial hardship.”³⁰⁵ The rationale underlying this exclusion is straightforward: Broadcasters often had to truncate CMI (for example, credits) to fit time constraints or to interrupt programs to air news segments, “thereby deleting part of the [interrupted broadcast’s] credits.”³⁰⁶ To “[accommodate] the need for accurate copyright management information to the realities of the broadcast . . . marketplace,” an exception was made.³⁰⁷

An early draft of § 1202, however, did not include this express exemption. Instead, a House committee’s analysis of the draft shows that broadcasters were excluded from the provision’s reach implicitly, because “ordinary and customary practices of broadcasters [and] inadvertent omission of credits from broadcasts of audiovisual works . . . are not made with knowledge that they

plaintiffs have argued that the violation occurred at the training stage. *See N.Y. Times*, 777 F. Supp. 3d at 314 (“The *Daily News* and [Center for Investigative Reporting] complaints, however, plausibly allege CMI removal during the LLM training process.”).

304. Plaintiffs thus find themselves resorting to back-of-the-napkin math to allege damages. *See, e.g.*, Second Amended Complaint, *supra* note 12, ¶ 103 (“In June 2022, Copilot had 1,200,000 users. If only 1% of users have ever received Output based on Licensed Materials and only once each, Defendants have ‘only’ breached Plaintiffs’ and the Class’s Licenses 12,000 times. However, each time Copilot outputs Licensed Materials without attribution, the copyright notice, or the License Terms it violates the DMCA three times. Thus, even using this extreme underestimate, Copilot has ‘only’ violated the DMCA 36,000 times.”).

305. 17 U.S.C. § 1202(e). Note that the exemption also requires that broadcasters “[do] not intend, by engaging in such activity, to induce, enable, facilitate, or conceal infringement of a right under this title.” *Id.*

306. *WIPO Treaties Hearing*, *supra* note 49, at 51 (statement of Marybeth Peters, Register of Copyrights).

307. GOLDSTEIN, *supra* note 102, § 7.18.

will induce, enable, facilitate or conceal a copyright infringement.”³⁰⁸ As discussed in Part I, the later inclusion of this carveout in the language of the provision suggests a desire to quell legal uncertainty rather than an implication that broadcasters are liable without the exception.³⁰⁹

This implied carveout creates a possibility of further exceptions without modern legislative intervention. Just as § 1202(e) exempts broadcasters where avoiding a violation is technically infeasible, courts should read § 1202’s scienter requirement to exempt LLM developers from liability when it is technically infeasible to avoid a violation—that is, when outputs are not identical to works used as training data.

While the legislative history surrounding § 1202 may be generally “conclusory and unilluminating,”³¹⁰ several features suggest the strength of Congress’s interest in an express broadcaster exemption. First, though the House Report on the DMCA read a broadcaster exception into the statute implicitly,³¹¹ the House Commerce Committee chose to codify it explicitly in § 1202(e).³¹² Second, this exemption was adopted despite its apparent conflict with the WIPO treaties the DMCA purported to enact, which specifically require signatories to hold broadcasters liable for enabling copyright violations.³¹³ Finally, given the extraordinary position of broadcasters to distribute large amounts of copyrighted work, the broadcaster exemption cuts strongly against what some might have thought was the spirit of the DMCA.³¹⁴ That Congress found the broadcaster exemption important is clear.

Like broadcasters, LLM developers face genuine technological limitations when it comes to preserving CMI for nonidentical outputs. Though research into novel attribution methods is ongoing,³¹⁵ these technologies are neither

308. H.R. REP. NO. 105-551, pt. 1, at 21 (1998).

309. *Supra* note 49 and accompanying text (explaining why such an exception might have been explicitly legislated).

310. David Nimmer, *Appreciating Legislative History: The Sweet and Sour Spots of the DMCA’s Commentary*, 23 CARDOZO L. REV. 909, 924 (2002).

311. *See* 17 U.S.C. § 1202(e); H.R. REP. NO. 105-551, pt. 1 at 20-21 (1998) (“The prohibition in this subsection does not include ordinary and customary practices of broadcasters or inadvertent omission of credits from broadcasts of audiovisual works since, inter alia, such omissions do not entail knowing provision of false CMI with intent to induce, enable, facilitate or conceal a copyright infringement.”).

312. *See* Nimmer, *supra* note 310, at 923. *Contrast* H.R. REP. NO. 105-551, pt. 1 at 39-40 (1998) (listing the proposed changes without § 1202(e)), *with* H.R. REP. NO. 105-796 at 15-16 (1998) (including § 1202(e)).

313. WIPO Copyright Treaty art. 12(1)(ii), Dec. 20, 1996, S. TREATY DOC. NO. 105-17, 2186 U.N.T.S. 121; WIPO Performances and Phonograms Treaty art. 19(1)(ii), S. TREATY DOC. NO. 105-17, 2186 U.N.T.S. 203.

314. Harbert, *supra* note 49, at 142.

315. *See, e.g.,* Wang et al., *supra* note 263, at 2.

finalized nor implemented, and requiring their use would impose significant financial burdens. But unlike broadcasters, LLMs *can* feasibly identify and filter for exact duplication,³¹⁶ meaning that under a technical feasibility standard they would remain liable for identical outputs. And unlike the broadcaster exemption, this proposed carveout would not conflict with international treaty obligations.³¹⁷ In fact, courts on one side of the current district court split have already adopted this approach in practice.

Intent and Damages.—The double scienter requirement was intended to cabin the otherwise broad § 1202 and justify its potential for extremely high statutory damages awards.³¹⁸ But it has not always done so, and it has created an opportunity for plaintiffs—especially in suits involving LLMs—to deter innovative and non-violative uses of works.³¹⁹ Where double scienter fails to limit the reach of § 1202, identity can step in as the provision’s limiting force by ensuring damages are awarded for only the most egregious and avoidable violations. Identity also simplifies the scienter analysis: Where outputs are nonidentical, courts need not analyze scienter at all, because no § 1202 violation can be alleged. Where outputs are identical, on the other hand, both plaintiffs and defendants are benefitted: Because identical works are more likely to be infringing compared to similar works,³²⁰ and because a defendant’s knowledge of their own infringement may satisfy the second scienter requirement,³²¹ it is relatively straightforward for plaintiffs to show that defendants “induce[d], enable[d], facilitate[d], or conceal[ed] an infringement.”³²² As for defendants, an identity requirement provides a realistic opportunity to avoid a claim altogether by using feasible technical capabilities to prevent identical outputs.

Copyrightability.—Though a leading treatise has suggested that liability under § 1202 in the context of thin copyright should require a showing approaching identity,³²³ courts applying § 1202 have not analyzed the

316. *See supra* notes 273-75 and accompanying text (describing LLM developers’ ability to check for identity).

317. *See supra* note 313 and accompanying text (explaining how the broadcaster exemption conflicted with World Intellectual Property Organization treaties).

318. Truskowski, *supra* note 224, at 1021-22 (referring to the double scienter requirement as one of the “internal safeguards” helping to ensure plaintiffs only recover against “knowing bad actors who remove or falsify CMI for their own gain”).

319. *See supra* text accompanying notes 243-53 (discussing the ease of alleging scienter where LLMs are at issue).

320. *See infra* notes 331-34 and accompanying text.

321. *See supra* note 220.

322. 17 U.S.C. § 1202(b).

323. 4 NIMMER & NIMMER, *supra* note 22, § 12A.10(C)(1).

copyrightability of a work's constituent parts.³²⁴ An identity requirement would restore balance by accounting for thin copyrightability without imposing an additional evidentiary burden on plaintiffs or courts. Identity makes a copyrightability analysis unnecessary: Even if the copyright in the original work is extremely thin, the allegedly violating work necessarily includes the copyrighted portions because it is identical to the original work. By contrast, a similarity standard would require courts to determine which elements are copyrightable through tools like the idea-expression dichotomy and the merger doctrine to ensure the elements being compared are not entirely unprotectable.³²⁵

The identity requirement has been critiqued for creating opportunities for defendants to skirt liability by outputting an exact copy of a plaintiff's work but changing, for example, just 1% to avoid identity.³²⁶ In the *GitHub* litigation, for instance, the plaintiff argued that Copilot's outputs differed from their source code only in ways that were "not semantically meaningful."³²⁷ But one scholar insists that those differences were "precisely the elements that . . . [were] most relevant to copyrightability."³²⁸ If indeed those overlapping elements were largely nonexpressive and, thus, unprotectable,³²⁹ then applying an identity standard rightfully protected GitHub from liability and helped ensure that copyright continued to protect expression, not ideas.³³⁰

Fair Use.—An identity requirement also mitigates the unfairness and perverse incentives created by § 1202's lack of a fair use defense. It narrows the scope of potential liability to the most egregious cases—those involving outright duplication—where fair use is least likely to apply anyway. At the same time, it prevents opportunistic plaintiffs from exploiting § 1202 to target defendants whose uses they cannot reach under traditional copyright doctrine due to the fair use exception.

324. See *supra* notes 138-39 and accompanying text.

325. These benefits are lost where an allegedly violated work is *entirely* uncopyrightable—for instance, if it is in the public domain. But in that instance, inability to satisfy a different element—the scienter requirement—would dispose of the claim.

326. See *infra* note 335 (describing this critique).

327. Complaint ¶ 60, *Doe v. GitHub, Inc.*, 672 F. Supp. 3d 837 (N.D. Cal. 2023) (No. 22-cv-06823), 2022 WL 16743590, ECF No. 1.

328. Duan, *supra* note 267, at 7.

329. Note that much of the allegedly infringing output material is redacted from the complaint. See Second Amended Complaint, *supra* note 12, ¶¶ 113-37.

330. See *Baker v. Selden*, 101 U.S. 99, 100-02 (1879); 17 U.S.C. § 102(b); see also Carys J. Craig, *Transforming "Total Concept and Feel": Dialogic Creativity and Copyright's Substantial Similarity Doctrine*, 38 CARDOZO ARTS & ENT. L.J. 603, 622 (2020) ("A substantial similarity doctrine that fails to parse out [unprotectable] elements from the protectable scope of the copyright work . . . is fundamentally at odds with a copyright system . . .").

Because of its overbreadth, § 1202 can tempt plaintiffs to bypass the limitations of a traditional copyright infringement claim—especially fair use. If a plaintiff anticipates that a fair use defense would defeat their infringement claim, they may instead pursue a § 1202 action based on the omission of CMI from a merely similar work. This tactic effectively converts § 1202 into a backdoor for claims that would otherwise be barred by fair use, expanding the statute beyond its intended scope.

Requiring identity as a threshold for § 1202 liability removes this incentive because fair use is less likely to apply to a work that is identical to a preexisting work. Two of the four factors used in the fair use analysis support this conclusion: the purpose or character of the use, and the amount and substantiality of the use.

The purpose or character of the use is the most important factor in the fair use analysis.³³¹ Identity can serve as “evidence of [the secondary work’s] lack of transformative purpose,” thus weighing against a finding of fair use.³³² For obvious reasons, reproducing a work identically also weighs against another factor: the amount and substantiality of the use.³³³ While fair use can sometimes apply to identical copying,³³⁴ it is far less likely to excuse an exact copy than to protect a transformative or merely similar work.

Under a § 1202 identity requirement, then, plaintiffs challenging merely similar uses will know that their claims cannot proceed. Plaintiffs targeting identical uses may still sue, but they are also less likely to be defeated in their copyright infringement claim by fair use in the first place and thus have less reason to use § 1202 as a workaround. In this way, identity curtails § 1202’s potential for misuse as a surrogate for nonviable infringement claims.

Addressing Gamesmanship.—A common critique of the identity requirement is that it may enable bad faith actors to evade liability by, for

331. See *Castle Rock Ent. Inc. v. Carol Publ’g Grp.*, 150 F.3d 132, 142 (2d Cir. 1998).

332. *Id.* at 143.

333. 4 NIMMER & NIMMER, *supra* note 22, § 13F.07 (“The third factor weighs against fair use when defendant copies a quantitatively large percentage of plaintiff’s work.”); *id.* (noting that “[a]s a rule, ‘the more of a copyright work that is taken, the less likely the use is to be fair,’” but that the infringer may still be favored “[i]f the amount copied is very slight in relation to the work as a whole” (second alteration in original) (first quoting *Swatch Grp. Mgmt. Servs. Ltd. v. Bloomberg L.P.*, 756 F.3d 73, 89 (2d Cir. 2014); and then quoting *Ringgold v. Black Ent. Television*, 126 F.3d 70, 76 (2d Cir. 1997))).

334. See *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1165 (9th Cir. 2007); 4 NIMMER & NIMMER, *supra* note 22, § 13F.05 (“Literal copying of images for a search engine, for example, has been held to satisfy the test because the defendant’s transformative purpose in using an image to facilitate search functionality is ‘fundamentally different’ from the original artist’s expressive purposes.” (emphasis omitted) (quoting *Perfect 10*, 508 F.3d at 1168)).

example, intentionally copying only 99% of a work.³³⁵ This concern is an important one; if it is true that the identity requirement could be skirted with a simple tweak—for example, to a single word in a manuscript—then LLM providers would effectively have a free pass to violate § 1202 through gamesmanship.

One option for addressing this concern lies in the language of § 1202(e)'s broadcaster exception. In addition to the requirement that the activity be technically or financially infeasible for the exception to apply, it requires that the would-be violator must not “intend, by engaging in such activity, to induce, enable, facilitate, or conceal infringement.”³³⁶ This requirement—which resembles the second scienter requirement—essentially imposes an obligation of good faith. While the second scienter requirement for removal or alteration of CMI requires *knowledge* (or constructive knowledge) that one is inducing, enabling, facilitating, or concealing an infringement,³³⁷ this condition of the broadcaster carveout prohibits *intention* of the same. This means broadcasters may enjoy an exemption from liability despite knowingly facilitating infringement so long as they do not intend to do so. Reading a good-faith obligation into the identity requirement would likewise prohibit LLM providers from intentionally facilitating infringement, thereby reducing the provision's vulnerability to gamesmanship through abuse of identity.

Though this particular good-faith obligation appeared for the first time when the broadcaster exception was translated from implicit to explicit via § 1202(e),³³⁸ an alternative basis for reading into the statute an *implicit* good-faith requirement is the § 1202(b) knowledge requirement. This is because “knowledge” under § 1202(b) can be read in one of two ways. First, there is the

335. *Real World Media LLC v. Daily Caller, Inc.*, 744 F. Supp. 3d 24, 40 (D.D.C. 2024) (“There is also a practical point: it 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.”); 4 NIMMER & NIMMER, *supra* note 22, § 12A.10(C)(1) (“If plaintiff owns the copyright to a 300-page book and defendant propounds a work in which a single sentence is missing from that work, the two are not *identical*—but are still beyond doubt *substantially similar*.”); Jane C. Ginsburg, *Humanist Copyright*, 6 J. FREE SPEECH L. 91, 128 (2025) (citing the 99% copying critique in stating that “one may hope the [identity] rule’s doctrinal and consequential flaws lead to its rejection”).

336. 17 U.S.C. § 1202(e)(1).

337. *Id.* § 1202(b).

338. *Contrast* H.R. REP. NO. 105-551, pt. 1, at 21 (1998) (“The prohibition in this subsection does not include ordinary and customary practices of broadcasters or inadvertent omission of credits from broadcasts of audiovisual works since, inter alia, such omissions are not made with *knowledge* that they will *induce, enable, facilitate or conceal a copyright infringement.*” (emphasis added)), with 17 U.S.C. § 1202(e)(1)(B) (exempting liability for broadcasters if “such person did not *intend, by engaging in such activity, to induce, enable, facilitate, or conceal infringement.*” (emphasis added)).

knowledge that inducement³³⁹ is happening as a matter of statistical certainty. That is, so long as LLMs retain a small chance of regurgitating their training data—and those LLMs are used to generate countless outputs—the developers can be almost certain that the LLMs they are producing are going to present copyrighted works to users without the associated CMI. Second, there is the direct personal knowledge that inducement is happening—that is, a person manually uploads a work without its CMI, or a social media platform publishes all photos uploaded by users after automatically stripping the photos' metadata.³⁴⁰

The House of Representatives' report on the implementation of the WIPO treaty suggests that the latter level of knowledge is required here. It explains that under the knowledge requirement of § 1202(b)—regardless of any exemption—broadcasters of audiovisual works do not violate the statute.³⁴¹ This suggests that when broadcasters accept that the variability in the length of a football game means that the credits of the following film might get cut off, they are not knowingly inducing infringement under § 1202(b). This does not obviate the need for the § 1202(e) exemption—broadcasters knowingly induce infringement if they directly and manually interrupt a movie's credits for a one-off emergency broadcast.³⁴² Unlike where the broadcasters accepted the possibility of inducing infringement, in making an emergency broadcast the company knew with certainty they would cause the movie to be presented absent controlling CMI.

If, then, personal knowledge is required under § 1202(b), the statute would effectively carry a good-faith requirement for LLM developers. If a developer creates a program whose purpose is to duplicate copyright works—or whose utility is meaningfully enhanced by its tendency to duplicate copyrighted works—that developer would have personal knowledge of their program presenting copied works to users. But if a developer creates a program which will, as a statistical matter, inevitably copy works and that developer makes reasonable efforts to prevent that eventuality, the developer will only have statistical knowledge of what will happen at the time the program is released to users.

339. Or enablement, facilitation, or concealment. See 17 U.S.C. § 1202(b).

340. Kayleigh Lemery, *DMCA's Double Scienier Requirement: Allowing Intermediary Software Companies to be Worry-Free in Making a Buck Off Another's Dime*, 28 MARQ. INTELL. PROP. & INNOVATION L. REV. 191, 197 (2024).

341. See H.R. REP. NO. 105-551, pt. 1, at 21 (1998).

342. *WIPO Treaties Hearing*, *supra* note 49, at 51 (statement of Marybeth Peters, Register of Copyrights) (“For example, a broadcaster who interrupts a broadcast of a motion picture for a news bulletin, thereby deleting part of the motion picture's credits, would not fall within the prohibition.”).

This argument is not without fault. Certainly, the latter developer has actual knowledge should they review user activity after the fact. But a broadcaster also possesses actual knowledge that they presented works without CMI after the film has aired. As such, this knowledge distinction seems to both match the report's explanation of § 1202(b) and offer the pragmatic appeal of requiring good faith. Taken together, the broadcaster exception and the statutory knowledge requirement provide a doctrinal foothold for construing an identity requirement to contain an implicit good-faith obligation.

However, for this solution to address the gamesmanship problem, courts must stop using the identity requirement as a precondition to liability and must instead view it as a component of scienter. Courts that follow the identity requirement currently treat identity as a filter: Where an output is not identical to an allegedly violated work, these courts immediately dismiss the case.³⁴³ This order of analysis would allow gamesmanship cases to escape scrutiny, because it creates a risk that cases may be dismissed before the court reaches an analysis of scienter. Instead, the identity requirement may be conceptualized as creating a presumption of the required scienter that may be rebutted—for example, by showing that a given output is identical to the plaintiff's work through coincidence, or through mutual copying of a third party. Conversely, a lack of identity can create a presumption that the required scienter does not exist—one that may be rebutted through a showing of bad faith.

Courts may hesitate to read in an obligation of good faith amid the provision's scarce legislative history³⁴⁴ or to break with past cases in analyzing identity along with scienter. But if the alternatives are to either restrict an industry because of its technological limitations or allow bad-faith developers to freely cause the exact harms contemplated by Congress, good-faith identity and scienter-based identity both offer paths forward that realize the DMCA's goals of encouraging industry and maintaining copyright in an increasingly technological world.

Conclusion

This Note has argued that § 1202 of the DMCA, while designed to modernize copyright law for the digital age, now risks becoming an overbroad and exploitable alternative to traditional infringement claims—especially in

343. *See, e.g., Doe v. GitHub, Inc.*, No. 22-cv-06823, 2024 WL 235217, at *9 (N.D. Cal. Jan. 22, 2024) (dismissing the § 1202 claims upon determining that the works were not identical); *Advanta-STAR Auto. Rsch. Corp. of Am. v. Search Optics, LLC*, 672 F. Supp. 3d 1035, 1057 (S.D. Cal. 2023) (same); *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-193, 2020 WL 5991503, at *7 (C.D. Cal. Aug. 14, 2020) (same).

344. Nimmer, *supra* note 310, at 924.

the context of generative AI. Unlike infringement, § 1202 imposes liability without requiring a showing of copyrightability, offers no fair use defense, and may permit even more disproportionate statutory damages than copyright infringement claims. Although the statute includes a scienter requirement, that requirement is so minimal that it fails to constrain the statute's reach in practice.

To address these concerns, this Note argues for adopting an identity requirement in § 1202 claims against LLMs. The urgency of this proposal is underscored by the Ninth Circuit's upcoming decision in *GitHub*, which will determine whether similarity alone suffices for § 1202 liability. That ruling will clarify the law within the influential Ninth Circuit and can resolve the structural mismatch between § 1202 and the technological realities of AI training.

This is not to suggest that courts offer the only remedy. In crafting § 1202, Congress affirmatively carved out the broadcaster exemption in § 1202(e). It recognized that leaving interpretation to the courts could have stifled an important industry. Depending on how the identity circuit split develops, Congress should create a similar explicit exception for LLM providers that codifies their exclusion from § 1202. Like broadcasters, LLM providers fail to meet the scienter requirement where outputs are similar but not identical to plaintiffs' works, given their technical inability to prevent such outputs.

But the proper interpretation of § 1202's scienter requirement should not be limited to industries with the power to compel specific congressional action. Nor are further industry-specific exemptions a good use of congressional resources. To stop § 1202 from ending the debate over LLMs' fair use before it begins, and to ensure the proper application of § 1202 across industries, the courts must adopt an identity requirement. The future of AI must be shaped by thoughtful policy; it is, after all, among the most important technologies of our generation.