The Pay-to-Train Paradigm: How Private Settlements Are Rewriting Generative AI Copyright Law

KEVIN AFIESH[†]

The legal landscape for generative artificial intelligence (AI) and copyright is now shaped more by private settlements that overwhelmingly favor copyright holders than by courts or legislators. As AI companies rely on copyrighted materials to train their models, content owners have filed lawsuits challenging these practices. However, instead of clarifying fair use limits in court, most disputes are resolved through private settlements. These agreements impose licensing fees, content restrictions, and compliance requirements, creating a shadow regulatory system that governs how AI companies access and use training data. Although courts and agencies have begun to weigh in, their interventions offer only partial guidance, highlighting the persistent uncertainty surrounding fair use in generative AI. This Note argues that settlements are not merely resolving disputes but actively reshaping AI copyright law by establishing an informal licensing system that significantly benefits copyright holders and dominant AI developers. In turn, this sidelines smaller AI companies and reduces competition. The result is a privatized framework where copyright holders dictate the terms of AI development while avoiding judicial and legislative oversight. This Note examines the motivations behind these settlements, the legal norms they create, and whether policymakers should intervene. It considers three potential paths forward: (1) Congress could codify AI training as fair use to protect innovation and disrupt the emerging pay-to-train model; (2) lawmakers could require transparency and opt-out mechanisms to balance copyright control with generative AI development; or (3) Congress could take no action, allowing private deals to continue shaping AI copyright law without intervention. The future of AI copyright law hinges on whether policymakers rebalance the regulatory environment or allow private agreements to prevail.

[†] J.D. Candidate 2026, University of California, College of the Law, San Francisco; Executive Communications Editor, *UC Law Journal*. I owe immense gratitude to Professor Ben Depoorter for his invaluable guidance and expertise that shaped this Note from its earliest stages. I am also sincerely thankful to Editor-in-Chief Skylar Schoemig, Executive Notes Editor Maddy Heldman, Executive Production Editor Miranda Sinsheimer, and the entire *UCLJ* Volume 77 team for their tireless dedication and thoughtful precision in elevating this Note and strengthening my voice as a writer. To my family, friends, mentors, and colleagues, thank you for your profound encouragement, wisdom, and inspiration. Each of you has left a lasting mark on my growth as a person and professional. My deepest appreciation goes to my dad, mom, and brother—Greg, Maya, and Chris Afiesh—for their unconditional love, endless support, and unwavering faith in me. You have always been my foundation and greatest source of strength. There is no mountain too high to climb with you in my corner, and for that I am forever grateful.

TABLE OF CONTENTS

INTRODUCTION

The emergence of artificial intelligence (AI) has resulted in a profound shift in copyright law, leaving questions unanswered by courts and legislatures. This Note argues that in the current legal void, copyright disputes involving generative AI are increasingly resolved through private settlements that favor licensing and control by rightsholders, primarily driven by defendants' risk aversion. This trend produces a shadow regulatory regime in which copyright law is shaped not by courts or Congress but by confidential agreements that dictate how AI developers access training data and manage legal exposure. As a result, private actors are defining the legal boundaries of AI development, often to the exclusion of smaller firms and without public oversight.

Generative AI models rely on massive datasets containing copyrighted material to develop their capabilities.[1] As these models are increasingly employed to generate content, copyright holders have pushed back, arguing that the unlicensed use of their works to train AI models constitutes infringement.[2] In response, AI developers invoke the fair use doctrine, claiming that training AI models with data that includes copyrighted material is a transformative use.[3] Yet, rather than litigating these unresolved questions, both sides have primarily turned to private settlements.[4]

These settlements are not just resolving disputes—they are defining them. Currently, the United States lacks comprehensive AI-specific copyright laws.[5] Courts are left to apply traditional copyright doctrines to determine whether using copyright-protected works to train AI systems constitutes infringement or fair use.[6] This legal uncertainty strongly incentivizes parties to settle rather than risk an unfavorable precedent.[7] Stakeholders negotiate in the shadow of potential litigation, often driven by fears of massive statutory damages and unpredictable rulings.[8] The recent decision in *Thomson Reuters v. Ross Intelligence*, which rejected a fair use defense in a non-generative AI context,[9] illustrates how courts

1. CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 4 (2025).

2. *Id.*; CITRIS AND THE BANATAO INSTITUTE, *Generative AI Meets Copyright–Pamela Samuelson*, at 05:30–07:41 (YouTube, Apr. 26, 2023), https://www.youtube.com/watch?v=6sDGIrVO6mo&t=2s.

3. CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 3–4 (2025); CITRIS AND THE BANATAO INSTITUTE, *supra* note 2, at 13:59–14:22; *see* OpenAI, Comment Regarding Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation, at 1, https://www.uspto.gov/sites/default/files/documents/OpenAI_RFC-84-FR-58141.pdf.

4. Aruni Soni, *AI Models Force Media Firms to Pick Licensing or Litigation (1)*, BLOOMBERG L. (Aug. 5, 2024, at 12:31 PT), https://news.bloomberglaw.com/ip-law/generative-ai-forces-media-firms-to-pick-licensing-or-litigation.

5. Pamela Samuelson, *Generative AI Meets Copyright: Ongoing Lawsuits Could Affect Everyone Who Uses Generative AI*, 381 SCI. 158, 158 (2023).

6. Katherine Lee, A. Feder Cooper & James Grimmelmann, *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain*, 72 J. COPYRIGHT SOC'Y 251, 407–08 (2025).

7. James Gibson, *Risk Aversion and Rights Accretion in Intellectual Property Law*, 116 YALE L.J. 882, 890 (2007).

8. *Id.* at 884; *see* Samuelson, *supra* note 5.

9. 765 F. Supp. 3d 382, 401 (D. Del. 2025).

may approach these issues, but it left unanswered how fair use applies to the training of generative AI models, whose outputs may be considered more transformative. Subsequent rulings in *Bartz v. Anthropic* and *Kadrey v. Meta* addressed that question directly,[10] but their holdings remain partial and incomplete, leaving the scope of fair use for generative AI unsettled.

Taken together, these dynamics have allowed settlements to increasingly shape AI copyright law outside the courtroom by establishing licensing fees, content restrictions, and other contractual terms as new norms.[11] Copyright holders seek to maintain continued control over their works by agreeing to license their data or instructing AI developers to restrict AI outputs—a method that AI firms comply with to avoid litigation.[12] This trend has led to a shadow regulatory system in which private agreements set AI copyright rules, not laws or court decisions.[13]

In the absence of legislative or judicial oversight, AI copyright law is being developed through behind-closed-doors negotiations[14] where rightsholders, wielding significant leverage, often dictate the terms.[15] Concurrently, it disadvantages smaller AI startups that cannot afford or comply with the costly terms of emerging settlement standards.[16] Over time, this privatized method of enforcing copyright could make licensing the norm, bypassing fair use protections[17] and restricting access to training data required for new AI development.[18]

---

10. Bartz v. Anthropic PBC, 787 F. Supp. 3d 1007, 1019 (N.D. Cal. 2025); Kadrey v. Meta Platforms, Inc., 788 F. Supp. 3d 1026, 1059 (N.D. Cal. 2025).

11. David M. McIntosh, Georgina Jones Suzuki & Yam Schaal, *AI and the Copyright Liability Overhang: A Brief Summary of the Current State of AI-Related Copyright Litigation*, ROPES & GRAY (Apr. 2, 2024), https://www.ropesgray.com/en/insights/alerts/2024/04/ai-and-the-copyright-liability-overhang-a-brief-summary-of-the-current-state-of-ai-related.

12. *See* Samuelson, *supra* note 5, at 160; Demand for Jury Trial ¶¶ 1,9, Getty Images (US), Inc. v. Stability AI, Inc., No. 1:23-CV-00135 (D. Del. filed Feb. 3, 2023) [hereinafter Getty Images Demand]; Class Action Complaint & Jury Trial Demanded ¶ 6, Authors Guild v. OpenAI Inc., No. 1:23-CV-08292 (S.D.N.Y. Sep. 19, 2023) [hereinafter Authors Guild Complaint & Demand].

13. *See* Gibson, *supra* note 7, at 884; Ben Depoorter, *Law in the Shadow of Bargaining: The Feedback Effect of Civil Settlements*, 95 CORNELL L. REV. 957, 974–76, 983 (2010).

14. Depoorter, *supra* note 13, at 983.

15. *See* Complaint & Demand for Jury Trial ¶ 83, Concord Music Grp., Inc. v. Anthropic PBC, No. 3:23-CV-01092 (M.D. Tenn. filed Oct. 18, 2023) [hereinafter Concord Music Group Complaint & Demand]; Blake Brittain, *Anthropic Reaches Deal on AI 'Guardrails' in Lawsuit Over Music Lyrics*, REUTERS (Jan. 3, 2025, at 10:56 PT), https://www.reuters.com/legal/litigation/anthropic-reaches-deal-ai-guardrails-lawsuit-over-music-lyrics-2025-01-03; *Associated Press, OpenAI Partner to Explore Generative AI Use in News*, REUTERS (July 13, 2023, at 10:05 PT) [hereinafter *Associated Press*, REUTERS], https://www.reuters.com/business/media-telecom/associated-press-openai-partner-explore-generative-ai-use-news-2023-07-13; Emma Roth, *OpenAI's DALL-E Will Train on Shutterstock's Library for Six More Years*, VERGE (July 11, 2023, at 14:47 PT), https://www.theverge.com/2023/7/11/23791528/openai-shutterstock-images-partnership.

16. U.S. COPYRIGHT OFF., M-363, IDENTIFYING THE ECONOMIC IMPLICATIONS OF ARTIFICIAL INTELLIGENCE FOR COPYRIGHT POLICY 42 (Brent A. Lutes ed. 2025).

17. *See* Gibson, *supra* note 7, at 895–97; Frank Pasquale & Haochen Sun, *Consent and Compensation: Resolving Generative AI's Copyright Crisis*, 110 VA. L. REV. ONLINE 207, 218 (2024).

18. U.S. COPYRIGHT OFF., M-363, IDENTIFYING THE ECONOMIC IMPLICATIONS OF ARTIFICIAL INTELLIGENCE FOR COPYRIGHT POLICY 53 (Brent A. Lutes ed. 2025).

This Note proceeds in four parts. Part I provides background on AI training methods, the evolving fair use doctrine, and the unresolved legal questions at the heart of recent lawsuits. Part II explores how settlements have become a dominant enforcement mechanism, functioning as a de facto regulation that benefits powerful rightsholders and establishes a pay-to-train licensing model. Part III examines three potential policy responses: (1) legally defining AI training as fair use, (2) creating a transparency and opt-out framework for copyright holders, and (3) permitting settlements to continue defining AI copyright law. Part IV concludes by summarizing the long-term implications of this privatized framework and the need for legislative intervention.

I. THE LEGAL AND INDUSTRY LANDSCAPE OF AI COPYRIGHT DISPUTES

A. FAIR USE, TRAINING DATA, AND THE CORE LEGAL QUESTIONS

Generative AI models rely on massive datasets to learn how to create text, images, and other content.[19] The AI training process begins by aggregating large quantities of data from varying sources such as public websites, licensed datasets, and private collections.[20] Throughout this training process, AI systems learn to mimic human-like responses and generate creative works, known as outputs.[21] However, these datasets often contain copyrighted material, such as books, articles, photographs, and other creative works, which poses significant legal and ethical challenges.[22]

The use of copyrighted material in AI training datasets has sparked legal debates about whether this constitutes infringement or can be considered fair use.[23] In U.S. copyright law, authors have exclusive rights to reproduce, distribute, and create derivative works from their original content.[24] However, AI companies argue that training on copyrighted works does not violate these rights if the use is transformative, while copyright holders contend that such use infringes their intellectual property.[25] The fair use doctrine permits limited use

19. Xiaoguang Tu, Zhi He, Yi Huang, Zhi-Hao Zhang, Ming Yang & Jian Zhao, *An Overview of Large AI Models and Their Applications*, 2 VISUAL INTEL., Dec. 27, 2024, at 1; *see* CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 4 (2025).

20. *See* Tu et al., *supra* note 19, at 1, 11; CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 4 (2025).

21. Tu et al., *supra* note 19, at 6; James Hutson, *The Evolving Role of Copyright Law in the Age of AI-Generated Works,* 2 J. DIGIT. TECH. & L. 886, 895 (2024).

22. Adam Buick, *Copyright and AI Training Data—Transparency to the Rescue?*, 20 J. INTELL. PROP. L. & PRAC. 182, 182, 184 (2024); CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 4 (2025).

23. *See* CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 4 (2023).

24. 17 U.S.C. § 106 (2024).

25. *See* CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 4 (2025); Edward Lee, *Prompting Progress: Authorship in the Age of AI*, 76 FLA. L. REV. 1445, 1451, 1497 (2024); OpenAI, *supra* note 3, at 4; Scott M. Douglass & Dominic Rota, *The Fast-Moving Race Between Gen-AI and Copyright Law*, BAKER DONELSON (July 10, 2024), https://www.bakerdonelson.com/

of copyrighted works without permission for specific purposes, such as "criticism, comment, news reporting, teaching . . . , scholarship, or research."[26] When determining fair use, courts evaluate four factors: the purpose and character of the use, the nature of the copyrighted work, the amount and substantiality of the portion used, and the effect on the potential market for the original work.[27] Under this doctrine, a use is considered "transformative" when it adds new expression, meaning, or purpose to the original work rather than merely superseding it as a substitute.[28]

Authors Guild v. Google provides an important precedent in assessing fair use in the context of large-scale digital reproduction. In 2004, Google launched its Google Books project, digitizing millions of books to create a searchable database.[29] The Authors Guild alleged that Google's mass digitization infringed their copyrights.[30] The Second Circuit ruled in favor of Google, finding the project constituted fair use.[31] The court determined that Google's database enabled the public to search for specific terms and excerpts rather than providing full-text access to copyrighted works, making Google's reproduction of the protected works transformative.[32] The court emphasized that such copying was permissible because it provided otherwise unavailable information about the original work without supplanting the expressive content itself.[33] Crucially, the court distinguished this informational purpose from merely creating derivative works, noting that the statutory right to prepare "transform[ed]" content does not extend beyond the scope of fair use when copying enables public knowledge rather than expressive substitution.[34] Thus, Google's use of the copyrighted works did not substitute for the market of the original works but instead provided a novel informational tool beneficial to the public.[35] The Supreme Court later denied certiorari, leaving the decision intact as a longstanding fair use precedent.[36]

AI copyright debates frequently cite *Authors Guild v. Google*, supporting the argument that large-scale digitization for indexing and search functionality

---

the-fast-moving-race-between-gen-ai-and-copyright-law; Megan K. Bannigan, Christopher S. Ford, Samuel J. Allaman & Abigail Liles, *An Early Win for Copyright Owners in AI Cases as Court Rejects Fair Use Defense*, DEBEVOISE & PLIMPTON (Feb. 14, 2025), https://www.debevoise.com/-/media/files/insights/publications/2025/02/an-early-win-for-copyright-owners-in-ai-cases-as.pdf.

26. 17 U.S.C. § 107 (2024).

27. *Id.*; OpenAI, *supra* note 3, at 4.

28. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994); Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508, 523–24 (2023).

29. Authors Guild v. Google, Inc., 804 F.3d 202, 208 (2d Cir. 2015).

30. *Id.* at 206–07.

31. *Id.* at 229.

32. *Id.* at 221–22.

33. *Id.* at 207.

34. *Id.* at 215–16.

35. *Id.* at 209, 222.

36. *Id.*, *cert. denied*, 578 U.S. 941 (2016).

may qualify as fair use.[37] AI developers contend that, like Google's searchable database, training AI models is transformative because it generates entirely new outputs rather than merely reproducing original copyrighted content directly.[38] Proponents of AI training further argue that using copyrighted material fosters technological progress and enables advancements in natural language processing and generative creativity.[39] They maintain that AI models do not copy works verbatim; instead, they use training data to develop broad capabilities, akin to how humans learn from reading books and engaging with media.[40]

However, the legal uncertainty surrounding AI training has led to conflicting interpretations of fair use. Opponents argue AI-generated works challenge core copyright doctrines, prompting courts to reconsider whether AI models should be subject to the same fair use principles as human creators.[41] Rightsholders stress that AI-generated content often closely resembles original works, diminishing the market value of copyrighted content and undermining the incentive to create.[42] They caution that allowing unrestricted use of copyrighted material could tilt copyright law unfairly in favor of AI developers and against content creators.[43]

Thomson Reuters v. Ross Intelligence offers the first judicial ruling addressing whether training AI models on copyrighted material qualifies as fair use.[44] Ross Intelligence, a legal research startup, used thousands of Westlaw headnotes and editorial summaries owned by Thomson Reuters to train a machine learning model intended to enhance legal search functionality.[45] Thomson Reuters filed suit, alleging Ross infringed on its copyrights by copying protected content without authorization.[46] Ross asserted a fair use defense, arguing that its use was transformative because it repurposed the material to improve legal research rather than merely summarize case law as the original headnotes were designed to do.[47] The court rejected Ross's fair use claim, finding that the use was not transformative and was highly commercial in

37. *See* Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc., 765 F. Supp. 3d 382, 399–400 (D. Del. 2025); Bartz v. Anthropic PBC, 787 F.Supp. 3d 1007, 1023 (N.D. Cal. 2025); Kadrey v. Meta Platforms, Inc., 788 F. Supp. 3d 1026, 1047, 1051 (N.D. Cal. 2025).

38. Buick, *supra* note 22, at 189.

39. Tu et al., *supra* note 19, at 1, 7, 12, 15; Buick, *supra* note 22, at 182; OpenAI, *supra* note 3, at 2.

40. *See* OpenAI, supra note 3, at 5; Jenny Quang, *Does Training AI Violate Copyright Law?*, 36 BERKELEY TECH. L.J. 1407, 1410 (2021).

41. *See* Mark A. Lemley, *How Generative AI Turns Copyright Upside Down*, 25 COLUM. SCI. & TECH. L. REV. 190, 191, 194 (2025); Hutson, *supra* note 21, at 896.

42. U.S. COPYRIGHT OFF., M-363, IDENTIFYING THE ECONOMIC IMPLICATIONS OF ARTIFICIAL INTELLIGENCE FOR COPYRIGHT POLICY 17 (Brent A. Lutes ed. 2025).

43. *Id.* at 40.

44. *See* Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc., 765 F. Supp. 3d 382, 397 (D. Del. 2025).

45. *Thomson Reuters*, 765 F. Supp. 3d at 390–91.

46. *Id.* at 391.

47. *Id.* at 398.

nature.[48] The court highlighted Ross's use of the expressive elements—not merely factual content—of Thomson Reuters' headnotes, and reasoned that this use posed a risk of market substitution.[49]

In rejecting Ross's defense, the court relied on recent Supreme Court guidance from *Warhol v. Goldsmith*, emphasizing that transformativeness requires a "further purpose or different character" from the original use.[50] Because both Ross and Westlaw used the headnotes to facilitate legal research, the court found the purposes too closely aligned to support a transformative use.[51] The court also dismissed Ross's claim that intermediate copying for AI training resembled transformative software cases like *Google v. Oracle*, noting that those involved computer code and interoperability, not expressive legal summaries.[52]

The ruling marked a significant development in the fair use debate over AI training, suggesting courts may be skeptical of unlicensed uses of copyrighted material for machine learning purposes. However, *Thomson Reuters* involved a non-generative AI system that did not produce new expressive outputs but rather improved legal search. Consequently, the decision left unresolved whether generative AI training capable of creating novel content would receive different legal treatment under the fair use doctrine. Two subsequent district court cases addressed this question directly and issued the first rulings on whether generative AI training can qualify as fair use.

The first, *Bartz v. Anthropic*, began in August 2024 when a group of authors filed a class-action lawsuit against Anthropic in the U.S. District Court for the Northern District of California.[53] The complaint alleged that Anthropic copied millions of books by scanning lawfully purchased print editions into digital form and by downloading unlawfully obtained digital files from shadow libraries such as Books3, Library Genesis, and Pirate Library Mirror.[54] Anthropic had assembled these materials into a centralized "research library," which served as the company's internal foundation for training its large language models, including its AI chatbot, Claude.[55] The plaintiffs claimed that both the storing of their works in Anthropic's library and the subsequent training of its large language models constituted copyright infringement.[56] Anthropic raised a fair use defense, arguing that both the ingestion of copyrighted works for training and the creation of the library were transformative uses.[57]

48. *Id.* at 397–398.
49. *Id.* at 400.
50. *Id.* at 397–98 (citation omitted).
51. *Id.* at 398.
52. *Id.*
53. 787 F. Supp. 3d 1007, 1018 (N.D. Cal. 2025).
54. *Id.* at 1014.
55. *Id.* at 1016.
56. *Id.* at 1014, 1018.
57. *Id.* at 1018.

Addressing Anthropic's defense, the court distinguished between works Anthropic obtained legally and those it copied from pirated sources.[58] It found that training Claude on plaintiffs' books could be "exceedingly transformative" when the inputs were lawfully acquired, reasoning that the process repurposed the works to teach the model linguistic patterns rather than to reproduce them.[59] The court further upheld Anthropic's scanning of its own purchased print books, treating the creation of digital copies for internal research as permissible.[60] But it declined to extend fair use to Anthropic's downloading and retention of pirated books, noting that building a permanent library from unlawful sources was itself infringing.[61] The court did not reach the question of whether training on those pirated works could qualify as fair use, leaving that issue to be addressed at a later stage of the trial.[62]

Before the issue could be resolved, however, Anthropic reached an unanticipated settlement with the plaintiffs in August 2025.[63] The agreement came just days before a scheduled hearing, abruptly ending the litigation and foreclosing appellate review.[64] By cutting the case short, the settlement left a highly consequential question—whether training on unlawfully obtained data can ever fall within fair use—unanswered. *Bartz* thus provided the first judicial holding on generative AI training under the fair use doctrine, but it left the doctrine incomplete.

The second major lawsuit, *Kadrey v. Meta*, commenced in July 2023 when a group of authors sued Meta Platforms in the U.S. District Court for the Northern District of California.[65] The complaint alleged that Meta reproduced the authors' books without authorization by downloading files from shadow libraries and using them to train its Llama large language models,[66] similar to the practices alleged against Anthropic in *Bartz*. The plaintiffs argued that Meta's acquisition of their books and their use in generative AI training constituted copyright infringement because the process enabled its models to generate outputs imitating the authors' style and content.[67] Meta raised a fair use defense, contending that the training process was transformative because its

58. *Id.* at 1032.

59. *Id.* at 1019, 1021–22.

60. *Id.* at 1023–24.

61. *Id.* at 1019.

62. *Id.* at 1034.

63. Order re Settlement in Principle, *Bartz*, 787 F. Supp. 3d 1007 (N.D. Cal. 2025) (No. 3:24-CV-05417); *see also* Brittain, *Anthropic Settles Class Action from US Authors Alleging Copyright Infringement*, REUTERS (Aug. 26, 2025, at 14:07 PT), https://www.reuters.com/sustainability/boards-policy-regulation/anthropic-settles-class-action-us-authors-alleging-copyright-infringement-2025-08-26.

64. Blake Brittain, *Anthropic's Surprise Settlement Adds New Wrinkle in AI Copyright War*, REUTERS (Aug. 28, 2025, at 08:37 PT), https://www.reuters.com/legal/government/anthropics-surprise-settlement-adds-new-wrinkle-ai-copyright-war-2025-08-27.

65. Kadrey v. Meta Platforms, Inc., 788 F. Supp. 3d 1026, 1036 (N.D. Cal. 2025).

66. Third Amended Consolidated Complaint ¶¶ 5–7, *Kadrey*, 788 F. Supp. 3d (N.D. Cal. 2025) (No. 3:23-CV-03417).

67. *Id.*

models did not reproduce the works directly but instead learned general stylistic patterns.[68]

In evaluating Meta's defense on the record before it, the court concluded that training the Llama models on the plaintiffs' books qualified as fair use.[69] The court explained that the process repurposed the works to extract statistical relationships rather than to reproduce their expressive content and thus did not substitute for the original books.[70] But the court emphasized that its holding turned on the plaintiffs' failure to present any meaningful evidence of market harm.[71] The judge acknowledged that a stronger evidentiary record could have compelled a different outcome, even suggesting that the plaintiffs might have prevailed on summary judgment had they made a sufficient showing.[72] As a result, the ruling provided only provisional support for treating generative training as transformative under fair use and left the doctrine unsettled.

The United States Copyright Office (USCO) similarly underscores the unsettled nature of this inquiry. The USCO's guidance acknowledges that training may appear transformative because models learn statistical patterns rather than reproduce works directly.[73] At the same time, the USCO emphasizes that the source of training data matters, cautioning that unlawful acquisition weighs heavily against a finding of fair use, especially where developers know or should know the works were unlawfully obtained.[74] The USCO further stresses that potential market harm—particularly the loss of licensing opportunities and the risk of market dilution from imitative outputs—must weigh heavily in evaluating the effect of training on the potential market for copyrighted works.[75] Yet the USCO has not taken a definitive position on whether generative AI training constitutes fair use, reinforcing that the question remains unresolved.[76]

Together, *Thomson Reuters*, *Bartz*, and *Kadrey* reflect a growing but fragmented body of case law on AI training. *Thomson Reuters* resolved a dispute over a non-generative system, but by doing so, it left open how the fair use doctrine should apply when models are trained to generate expressive outputs. *Bartz* provided the first recognition that training on lawfully obtained materials can qualify as fair use, yet the court withheld that protection for unlawfully sourced data, leaving the doctrine incomplete on a central question. *Kadrey* sustained generative training as fair use on the record before it, but its reasoning suggested the outcome could differ in future cases where plaintiffs substantiate

68. *Kadrey*, 788 F. Supp at 1047–48.
69. *Id.*
70. *Id.* at 1039.
71. *Id.* at 1060.
72. *Id.*
73. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 45 (2025).
74. *Id.* at 45, 107.
75. *Id.* at 61.
76. *Id.* at 74.

claims of market harm. In addition, the USCO has stopped short of resolving whether generative training qualifies as fair use, cautioning only that transformativeness must be weighed against potential market harm. Taken together, these judicial and administrative signals offer only partial guidance, underscoring that the doctrinal boundaries of fair use in the context of generative AI remain nascent.

B. INDUSTRY ENFORCEMENT STRATEGIES IN A LEGAL VACUUM

Rightsholders have pursued copyright enforcement against generative AI developers with distinct strategies reflecting their respective industries. While litigation remains a tool to assert rights, the primary goal is often not acquiring damages but securing licensing terms to control developer access to copyrighted content. This shift has produced a fragmented but revealing picture of how different sectors respond to an unsettled legal terrain.

1. News Media—Litigation Pressure vs. Proactive Licensing

The news media have become a central battleground in the copyright debate, with leading publishers responding to AI's use of journalistic content. In *The New York Times v. Microsoft*, The New York Times filed a lawsuit in December 2023 in the U.S. District Court for the Southern District of New York against OpenAI and its primary investor, Microsoft, becoming the first major news organization to sue an AI powerhouse.[77] The New York Times alleges that OpenAI scraped millions of its articles without authorization to train ChatGPT, occasionally enabling the AI model to generate text containing verbatim excerpts from Times content.[78] The New York Times argues that this practice infringes its copyrights and threatens its subscription model by allowing users free access to information derived from its reporting.[79]

This lawsuit has attracted significant attention as it directly addresses whether AI companies should be required to license news media to train their AI models. OpenAI has defended its practices by asserting a fair use defense, contending its models transform input data rather than merely copying it.[80] However, The New York Times argues that ChatGPT's capacity to reproduce substantial portions of its content nearly verbatim weakens OpenAI's claim that the use is transformative.[81]

77. Complaint at 1, 68, N.Y. Times Co. v. Microsoft Corp., No. 1:23-CV-11195 (S.D.N.Y. 2023) [hereinafter N.Y. Times Co. Complaint]. This action has been consolidated for pretrial purposes in *In re OpenAI, Inc. Copyright Infringement Litigation*, No. 1:25-MD-03143 (S.D.N.Y. 2025), pursuant to the Judicial Panel on Multidistrict Litigation's Transfer Order. Sara Fischer, *NYT Sues OpenAI, Microsoft for Copyright Infringement*, AXIOS (Dec. 27, 2023), https://www.axios.com/2023/12/27/nyt-microsoft-openai-lawsuit-copyright-infringement.

78. N.Y. Times Co. Complaint, *supra* note 77, ¶ 160.

79. *Id.* ¶ 2.

80. *Id.* ¶¶ 7–8.

81. *Id.* ¶ 8.

Similarly in *Dow Jones v. Perplexity AI*, Dow Jones, publisher of The Wall Street Journal, sued Perplexity AI in October 2024 in the U.S. District Court for the Southern District of New York, alleging "massive illegal copying" of its news articles.[82] Dow Jones claims that Perplexity AI unlawfully scraped and republished its proprietary journalistic content to train its AI-powered search engine that directly answers user queries using material derived from The Wall Street Journal.[83] Unlike OpenAI, which has defended itself against The New York Times lawsuit on the grounds of fair use, Perplexity AI faces allegations of directly reproducing substantial portions of The Wall Street Journal articles without proper licensing or attribution.[84] Dow Jones argues that Perplexity AI's search engine effectively substitutes for a paid news subscription, thereby diverting potential revenue from traditional publishers.[85]

Not all publishers have followed The New York Times's and Dow Jones's approaches, however. In July 2023 and May 2024, respectively, the Associated Press and News Corp struck licensing deals with OpenAI, granting access to their vast news archives for AI training purposes.[86] These proactive agreements demonstrate that some publishers find collaborative licensing more pragmatic than litigation.[87]

The divergence in enforcement strategies stems from the economic structure of the news industry and the nature of the underlying harm. Generative AI typically produces summaries rather than reproducing entire articles, diverting readers from original platforms and reducing subscription-based and ad-supported revenue.[88] For publishers with substantial proprietary archives and sufficient resources, such as The New York Times and Dow Jones, litigation is a strategic opportunity to clarify fair use boundaries and leverage settlements or precedents.[89] Organizations with narrower margins or established syndication models find direct negotiations with AI companies more practical to secure compensation and influence evolving norms without litigation.[90]

82. Complaint ¶ 77, Dow Jones & Co. v. Perplexity AI, Inc., No. 24-CV-07984 (S.D.N.Y. Oct. 21, 2024) [hereinafter Dow Jones Complaint].

83. *Id.* ¶¶ 53–54.

84. *Id.* ¶ 5.

85. *Id.* ¶¶ 1, 4.

86. Matt O'Brien, *ChatGPT-Maker OpenAI Signs Deal with AP to License News Stories*, AP NEWS (July 13, 2023, at 08:41 PT), https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a; Press Release, News Corp, News Corp and OpenAI Sign Landmark Multi-Year Global Partnership (May 22, 2024), https://investors.newscorp.com/node/15946/pdf.

87. *Id.*

88. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 31, 63–64, 63 n. 36 (2025).

89. *See* News/Media Alliance, Comment Letter on Notice of Inquiry and Request for Comments on Study of Copyright Law and Artificial Intelligence 16 (Oct. 30, 2023), https://www.newsmediaalliance.org/wp-content/uploads/2023/10/Final-NMA-USCO-AI-NOI-Response-Submitted-10.30.23.pdf.

90. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 67 (2025).

The outcomes of *The New York Times* and *Dow Jones* cases will significantly shape future interactions between AI companies and news publishers, particularly regarding whether licensing becomes the industry standard for training data. Nevertheless, the strategic split within the industry reflects a broader debate of whether publishers should litigate to defend their rights or collaborate with AI developers to influence the standards that will govern the industry.

2. Publishing and Literature—Familiar Choices, Diverging Paths

The publishing industry has similarly initiated legal action against AI companies, with *Authors Guild v. OpenAI* emerging as a notable test case. In September 2023, the Authors Guild—representing prominent authors such as John Grisham, George R.R. Martin, and Jodi Picoult—filed a class-action lawsuit against OpenAI in the U.S. District Court for the Southern District of New York.[91] The lawsuit alleges OpenAI unlawfully copied copyrighted books to train its AI models, enabling those models to generate text that closely imitates the style and content of the original authors.[92]

This case draws comparisons to *Authors Guild v. Google*, where Google successfully defended its mass digitization of books as fair use.[93] However, the Authors Guild now argues that OpenAI's use goes further, creating outputs that can directly substitute authors' original works.[94] OpenAI maintains that training generative AI models falls within fair use and is essential to technological advancements in generative AI.[95]

As the lawsuit continues, some publishers are pursuing alternative strategies. In late 2023, HarperCollins Publishers entered into a licensing agreement with Microsoft, granting the tech giant access to its catalog to train its generative AI tools.[96] Around the same time, Cambridge University Press, a major academic publisher, publicly acknowledged that it had begun engaging with AI developers regarding licensing deals.[97] Additionally, Cambridge University Press reached out to its authors to amend existing publication

91. Authors Guild Complaint & Demand, *supra* note 12, ¶¶ 1, 11–28. This action has been consolidated with other class action lawsuits against OpenAI for pretrial purposes. Transfer Order, *In re* OpenAI, Inc., Copyright Infringement Litig., No. 25-MD-03143 (S.D.N.Y. Apr. 3, 2025).

92. Authors Guild Complaint & Demand , *supra* note 12, ¶¶ 1–3.

93. Authors Guild v. Google, Inc., 804 F.3d 202, 229 (2d Cir. 2015).

94. Authors Guild Complaint & Demand , *supra* note 12, ¶ 2; *see* Micaela Mantegna, *ARTificial: Why Copyright Is Not the Right Policy Tool to Deal with Generative AI*, 133 YALE L.J. F. 1126, 1154 (2024).

95. Authors Guild Complaint & Demand, *supra* note 12, ¶¶ 109–10.

96. Hannah Miller & Dina Bass, *Microsoft Signs AI-Learning Deal with News Corp.'s HarperCollins*, BL (Nov. 19, 2024, at 09:44 PT), https://www.bloomberglaw.com/product/blaw/bloomberglawnews/bloomberg-law-news/X1NLS5UO000000.

97. *Uncharted Territory: AI and the Cambridge Approach for Academic Book Publishing*, CAMBRIDGE UNIV. PRESS (Sep. 13, 2024), https://www.cambridge.org/us/universitypress/about-us/news-and-blogs/uncharted-territory-ai-and-the-cambridge-approach-for-academic-book-publishing.

agreements, further suggesting a shift toward future licensing arrangements and encouraging their adoption among the publishing and literature industries.[98]

The mixed enforcement strategies in the publishing sector reflect differences in how rights are held and how content is circulated. Author organizations such as the Authors Guild have turned to litigation, arguing that AI models unlawfully ingest and reproduce expressive and stylistic elements of copyrighted books without consent.[99] These organizations are well-positioned to bring coordinated lawsuits on behalf of multiple authors and view litigation as a way to clarify the boundaries of fair use. Meanwhile, large commercial publishers like HarperCollins and Cambridge University Press retain centralized control over licensing rights and have opted to negotiate directly with AI developers without litigation.[100] These licensing decisions aim to mitigate concerns about diminished readership and reduced control over how published works are repurposed in AI-generated outputs.[101] As in the news media sector, these varied incentives and institutional structures have produced a dual-track enforcement strategy.

The outcome of *OpenAI* could still play a key role in shaping the legal boundaries of AI training within the publishing industry. A ruling against OpenAI may reinforce the view that training AI models on copyrighted books without authorization constitutes infringement, potentially accelerating the trend toward formal licensing agreements as seen with HarperCollins and Cambridge University Press. However, a ruling in OpenAI's favor could build upon the *Authors Guild v. Google* ruling and strengthen fair use defenses, offering broader legal coverage for AI developers. Regardless of the verdict, recent industry behavior suggests that litigation and licensing are evolving in tandem—both shaping how copyright law adapts to generative AI.

3. Visual Arts and Photography—Litigation or Collaboration

In February 2023, Getty Images filed suit against Stability AI, the developer of Stable Diffusion, in the U.S. District Court for the District of Delaware.[102] Getty Images alleged that Stability AI unlawfully scraped over twelve million copyrighted photographs from Getty's database to train its AI models.[103] Stability AI's models not only used these images without permission but also produced outputs closely resembling copyrighted works.[104] Getty

98. *Id.*

99. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 44,49 (2025).

100. Miller & Bass, *supra* note 96; *Uncharted Territory: AI and the Cambridge Approach for Academic Book Publishing*, *supra* note 97.

101. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 73, 89, 105 (2025); News/Media Alliance, *supra* note 89, at 4, 6, 8–9, 22.

102. Getty Images Demand, *supra* note 12, ¶¶ 1, 7.

103. *Id.* ¶ 1.

104. *Id.* ¶ 36.

further emphasized that Stable Diffusion occasionally generated outputs containing distorted versions of the Getty watermark, reinforcing its claim that Stability AI relied extensively on Getty's content.[105]

Getty Images v. Stability AI highlights the visual arts industry's willingness to take an aggressive stance when seeking to include licensing expectations for generative AI training. Getty has taken a more forceful legal approach, claiming Stability AI's conduct constitutes widespread copyright infringement, similar to the approach The New York Times and Dow Jones took in their respective lawsuits.[106] Getty is seeking monetary damages and injunctive relief, which could potentially compel Stability AI to destroy any models trained on Getty's images, including Stable Diffusion.[107] Stability AI has denied liability and has moved to dismiss the lawsuit, arguing that claims of infringing Getty's copyrights are too "vague and general."[108] *Getty Images* remains unresolved, but its outcome could significantly influence how AI companies interact with stock photography and visual arts.

Getty's legal strategy differs from that of other visual content providers who have chosen voluntary measures or licensing deals to navigate AI-generated content issues. For instance, Shutterstock, once a direct competitor of Getty,[109] adopted a cooperative strategy in July 2023 by entering into a licensing partnership with OpenAI.[110] Under this agreement, Shutterstock granted OpenAI authorized access to its extensive image database for AI training while also establishing a compensation mechanism for artists whose works contribute to the resulting AI-generated content.[111] Shutterstock's proactive licensing approach underscores a willingness to collaboratively address the complexities posed by AI, positioning itself favorably within evolving industry norms without relying on costly and adversarial litigation.

The visual arts sector, while still engaged in active litigation, appears to be trending toward licensing as the more sustainable enforcement strategy. Stock photography platforms like Shutterstock and Getty maintain centralized control over vast image libraries and possess the technical infrastructure, such as

105. *Id.* ¶ 11.

106. *Id.* ¶ 1; *see* N.Y. Times Co. Complaint, *supra* note 77, ¶¶ 2–3; Dow Jones Complaint, *supra* note 82, ¶¶ 2–3.

107. Getty Images Demand, *supra* note 12, ¶ 2.

108. Katie Prescott, *Stability AI Rejects Photo Copyright Claims in High Court Case*, TIMES (Feb. 29, 2024, at 18:40 GMT), https://www.thetimes.com/business-money/article/stability-ai-rejects-photo-copyright-claims-in-high-court-case-m63ln9mbl.

109. Getty Images and Shutterstock, once direct competitors in the stock photography market, announced in January 2025 that they had agreed to merge, marking a significant consolidation within the visual content licensing industry. Press Release, Shutterstock, Getty Images and Shutterstock to Merge, Creating a Premier Visual Content Company (Jan. 7, 2025), https://investor.shutterstock.com/static-files/d51b865e-a4ec-402e-8a1d-265f402a9079; Press Release, Getty Images, Getty Images and Shutterstock to Merge, Creating a Premier Visual Content Company (Jan. 7, 2025), https://investors.gettyimages.com/node/14036/pdf.

110. Press Release, Shutterstock, Shutterstock Expands Partnership with OpenAI, Signs New Six-Year Agreement to Provide High-Quality Training Data (July 11, 2023, at 13:00 ET), https://investor.shutterstock.com/news-releases/news-release-details/shutterstock-expands-partnership-openai-signs-new-six-year.

111. *Id.*

watermarking and metadata tagging, to monitor and monetize their content effectively.[112] This centralized rights management makes licensing easier to implement than in sectors like publishing, where rights are more fragmented and use cases more varied.[113] As AI models increasingly rely on large-scale datasets for training, visual content providers are well-positioned to negotiate licensing terms that provide compensation while reducing the risk of infringement.

Whether litigation or licensing will ultimately serve as the leading strategy for addressing generative AI disputes within the visual arts sector remains unclear. Regardless of which strategy prevails, high-profile disputes like *Getty Images* are poised to influence broader standards governing how AI companies may utilize copyrighted visual content.

4. Music Publishing—A Unified Front Against Generative AI

Unlike other creative sectors that have shown a split between litigation and its alternatives, the music industry has broadly adopted a unified stance of prioritizing content-based restrictions to prevent AI models from generating infringing outputs, as exemplified by *Concord Music Group v. Anthropic*.[114] In October 2023, a coalition of major music publishers—including Universal Music Group, Concord Music Group, and ABKCO Records—filed a lawsuit against Anthropic in the U.S. District Court for the Middle District of Tennessee.[115] The plaintiffs allege that Anthropic used copyrighted song lyrics without authorization to train its Claude AI chatbot, resulting in outputs that reproduce verbatim lyrics from songs such as Katy Perry's "Roar" and the Rolling Stones' "You Can't Always Get What You Want."[116] Although the lawsuit remains ongoing, Anthropic agreed to implement guardrails to prevent Claude from generating copyrighted lyrics, signaling a partial dispute resolution.[117]

Rather than focusing on licensing fees or negotiated compensation, the coalition of plaintiffs in *Anthropic* reflects an unusually unified stance

112. McGucken v. Shutterstock, Inc. No. 23-7652, 2026 WL 364412, at *4 (2d Cir. Feb. 10, 2026); Getty Images Demand, *supra* note 12, ¶¶ 4, 31, 57, 75; *see generally* Prasanth Vaidya Sanivarapu, Kandala N. V. P. S. Rajesh, Khalid M. Hosny & Mostafa M. Fouda, *Digital Watermarking System for Copyright Protection and Authentication of Images Using Cryptographic Techniques*, 12 APPLIED SCI. 8724, 8724 (2022), https://doi.org/10.3390/app12178724 (explaining that digital watermarking embeds copyright data into media to track use and protect against unauthorized distribution).

113. *Copyright and Generative AI: Recent Developments on the Use of Copyrighted Works in AI*, MCGUIREWOODS (Sep. 2, 2025), https://www.mcguirewoods.com/client-resources/alerts/2025/9/copyright-and-generative-ai-recent-developments-on-the-use-of-copyrighted-works-in-ai.

114. *See* Concord Music Group Complaint & Demand, *supra* note 15, ¶¶ 1–7.

115. *Id.* ¶¶ 1–2.

116. *Id.* ¶¶ 4, 66, 69.

117. Blake Brittain, *Anthropic Reaches Deal on AI 'Guardrails' in Lawsuit Over Music Lyrics*, REUTERS (Jan. 3, 2025 10:56 PT), https://www.reuters.com/legal/litigation/anthropic-reaches-deal-ai-guardrails-lawsuit-over-music-lyrics-2025-01-03.

prioritizing content-based restrictions on AI-generated outputs.[118] Driven by concerns that AI-generated music could directly substitute for original compositions, publishers have shifted focus from compensation toward maintaining control over their works.[119] This strict enforcement stance reflects both the structural and economic characteristics unique to music publishing.

Whereas news articles or books are longer-form and contextually adaptable, musical works—particularly lyrics and melodies—are short, recognizable, and easily replicated verbatim by generative AI with minimal prompting. This makes AI-generated musical content especially vulnerable to infringement.[120] When generative AI models produce recognizable lyrics or musical phrases, they threaten licensing revenue and royalty structures integral to the industry's economic model, potentially displacing musicians.[121] Additionally, the centralized licensing infrastructure and collective rights organizations that dominate music publishing offer fewer incentives to negotiate bespoke deals for training data.[122] These combined factors of high substitutability, stronger infringement claims, and an entrenched licensing infrastructure help explain why litigation has emerged as the dominant tool for enforcement in the music sector.

Collectively, the responses from the news, publishing, visual arts, and music industries highlight a critical imbalance in negotiating power between copyright holders and AI developers. While a licensing trend is emerging in the visual arts sector and hybrid strategies have taken hold in publishing and news media, music publishers have remained firmly aligned around litigation and output restrictions as their primary means of enforcement.[123] These differing approaches reflect the distinct economic and structural concerns that shape enforcement strategies across creative sectors.

Despite these industry-specific differences, a broader trend has emerged in which copyright holders are not waiting for courts or legislatures to define the legal status of AI training. Copyright holders have established themselves as de facto regulators, reshaping the legal and commercial landscape for AI development.[124] This evolving framework raises critical questions about the appropriate role of government and public institutions in establishing consistent and transparent rules.

---

118. *See* Concord Music Group Complaint & Demand, *supra* note 15, ¶¶ 1, 14; *see also* U.S. COPYRIGHT OFF., M-363, IDENTIFYING THE ECONOMIC IMPLICATIONS OF ARTIFICIAL INTELLIGENCE FOR COPYRIGHT POLICY 5 (Brent A. Lutes ed. 2025).

119. Concord Music Group Complaint & Demand, *supra* note 15, ¶¶ 47–48.

120. *See* U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 6–7 (2025).

121. *See id.* at 62.

122. *See* U.S. COPYRIGHT OFF., COPYRIGHT AND THE MUSIC MARKETPLACE 8 (2015).

123. *See* discussion *infra* Subpart.II.B.1–4.

124. *See* N.Y. Times Co. Complaint, *supra* note 77, ¶¶ 198–204; Dow Jones Complaint, *supra* note 82, ¶ 1; Authors Guild Complaint & Demand, *supra* note 12, ¶¶ 46–47; Getty Images Demand, *supra* note 12, ¶¶ 34–36; Concord Music Group Complaint & Demand, *supra* note 15, ¶¶ 57–59.

C. GOVERNMENT INTERVENTION AND EARLY REGULATORY EFFORTS

Despite AI's rapid adoption and groundbreaking development, no definitive laws currently govern the use of copyrighted works to train AI models. The USCO has issued guidance stating that AI-generated works do not qualify for copyright protection due to a lack of human authorship.[125] However, this guidance does not resolve whether using copyrighted works in AI training constitutes infringement or fair use. In response to increasing litigation and uncertainty, the USCO and the U.S. Patent and Trademark Office (USPTO) have launched public consultations to assess whether AI-generated content should be subject to new copyright frameworks.[126]

These agencies have actively sought input from various stakeholders, including AI companies, copyright holders, and policymakers, to address the existing legislative ambiguity surrounding AI copyright liability.[127] The USCO has published multiple reports examining AI's impact on current copyright law.[128] The reports emphasize that existing statutes do not adequately address the complexities introduced by generative AI, having been enacted without anticipating such technological advancements.[129] The reports further acknowledge that AI's ability to ingest and reproduce copyrighted material challenges traditional copyright enforcement, leading to increased pressure for legislative reform.[130]

Given these challenges, the USCO has considered several possible licensing frameworks for AI training, including voluntary licensing, compulsory licensing, extended collective licensing, and opt-out mechanisms, each with distinct tradeoffs for innovation and copyright protection.[131] Meanwhile, the recent decision in *Thomson Reuters* signals that AI training on copyrighted data may not fall under traditional fair use protections—at least in cases involving non-generative systems.[132] Nonetheless, courts have yet to definitively rule on whether the training of generative AI models qualifies as fair use.

Congress has also held multiple hearings on AI copyright liability, focusing on whether AI-generated content should be subject to existing copyright laws or

125. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 2: COPYRIGHTABILITY 8 (2025); *Managing AI Generated Content: Legal & Ethical Complexities*, LUMENOVA (June 18, 2024), https://www.lumenova.ai/blog/aigc-legal-ethical-complexities.

126. *See* Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190, 16191 (Mar. 16, 2023) (to be codified at 37 C.F.R. pt. 202).

127. U.S. COPYRIGHT OFF., M-363, IDENTIFYING THE ECONOMIC IMPLICATIONS OF ARTIFICIAL INTELLIGENCE FOR COPYRIGHT POLICY 5, 9 (Brent A. Lutes ed. 2025).

128. *Id.* at 9.

129. *Id.* at 3.

130. *Id.* at 31.

131. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 85, 95, 99, 101(2025).

132. Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc., 765 F. Supp. 3d 382, 401 (D. Del. 2025).

require a new legislative framework.[133] Testimony from legal scholars and industry representatives has garnered diverse viewpoints on whether AI companies should be required to obtain explicit licenses for training data or whether fair use should extend to AI training models.[134] Some lawmakers have proposed mandatory licensing schemes to address concerns from rightsholders, while others argue that overregulation could disrupt innovation.[135]

These regulatory discussions sharply contrast with current practice, where private settlements increasingly dictate AI copyright norms before lawmakers intervene. As novel AI copyright claims emerge, companies resolve disputes through confidential licensing agreements and out-of-court settlements. This dynamic raises critical concerns regarding whether private agreements will preempt judicial and legislative roles in defining broader copyright protections in AI. Ultimately, it remains to be seen whether Congress will intervene to formalize AI copyright regulations or whether industry-led settlements will continue to shape norms without explicit statutory guidance.

II. PRIVATE SETTLEMENTS AS THE DOMINANT FORCE IN AI COPYRIGHT LAW

A. WHY AI COMPANIES SETTLE INSTEAD OF LITIGATE

AI companies frequently opt into settlement agreements in copyright disputes rather than risk litigation due to legal uncertainty, financial considerations, and reputational concerns.[136] Unlike traditional copyright cases, disputes involving AI training raise unresolved legal questions, notably whether unlicensed use constitutes fair use or infringement.[137] Moreover, AI firms also prefer to preserve business relationships and maintain a positive public image, further making settlement agreements more attractive than litigation.[138]

1. Legal Risk, Financial Pressure, and Strategic Uncertainty

Courts have yet to conclusively resolve whether ingesting copyrighted works to train generative AI models constitutes transformative fair use or infringes upon reproduction and derivative rights.[139] While AI companies often analogize this practice to Google's digitization of books for search indexing—

133. CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 6 (2025).

134. *Id.*; Samuelson, *supra* note 5, at 160.

135. CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 6 (2025).

136. J.J. Prescott & Kathryn E. Spier, *A Comprehensive Theory of Civil Settlement*, 91 N.Y.U. L. REV. 59, 71, 99 (2016); s*ee* discussion *infra* Subpart.II.A.1.

137. Quang, *supra* note 40, at 1407–08.

138. *See* discussion *infra* Subpart.II.A.1.

139. CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 3 (2025); Gibson, *supra* note 7, at 889; Samuelson, *supra* note 5, at 159; Quang, *supra* note 40, at 1407–08.

previously deemed fair use—the distinction between indexing and generating original outputs complicates this argument.[140] Because AI-generated outputs may more directly substitute for original works, they face greater legal vulnerability than search functionalities.[141] This uncertainty strongly incentivizes settlement.

In addition to doctrinal ambiguity, AI companies bargain in the shadow of the law, where the risk of unfavorable precedent significantly shapes settlement strategies.[142] In high-stakes disputes, the potential systemic cost of an adverse ruling may outweigh any benefit from judicial clarity, particularly where doctrinal gaps increase the perceived volatility of outcomes.[143] The incentive to settle is further magnified when institutional players face repeat exposure and seek to avoid embedding risk into future legal environments.[144] In areas lacking clear judicial guidance, powerful private actors negotiate settlements to avoid court rulings that could undermine their legal positions.[145]

This bargaining process reflects what scholars have described as a game-theoretic environment in which parties, faced with unclear legal standards, calibrate their strategies based on anticipated court behavior and the opposing party's risk tolerance.[146] In such settings, legal uncertainty itself becomes a factor in negotiation, shaping whether disputes are settled or litigated before any judicial resolution occurs.[147]

This pattern reflects how litigation filters for only the most uncertain disputes—cases where both sides believe they might win, yet neither is willing to risk an adverse ruling that could shape legal doctrine.[148] Such settlements are particularly relevant in AI copyright disputes, serving as tools to preempt judicial outcomes that could threaten business models and long-term growth. Rightsholders, in turn, leverage this uncertainty to push companies into licensing frameworks before courts can codify a broad fair use defense.[149]

The economic stakes further compound this pressure to settle. The threat of statutory damages further incentivizes early settlement. Under 17 U.S.C. § 504(c)(2), if a court finds infringement was willful, it can award a copyright

140. *See* Samuelson, *supra* note 5, at 160.

141. *See id.* at 159; Hutson, *supra* note 21, at 896.

142. Gibson, *supra* note 7, at 887–88, 945–46; Ryan Abbott & Elizabeth Rothman, *Disrupting Creativity: Copyright Law in the Age of Generative Artifical Intelligence*, 75 FLA. L. REV. 1141, 1160 (2023); Robert Cooter, Stephen G. Marks & Robert Mnookin, *Bargaining in the Shadow of the Law: A Testable Model of Strategic Behavior*, 11 J. LEGAL STUD. 225, 225 (1982); *see* Prescott & Spier, *supra* note 136, at 127; George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1, 16 (1984).

143. Cooter et al., *supra* note 142, at 238; Prescott & Spier, *supra* note 136, at 60, 69; Priest & Klein, *supra* note 142.

144. Cooter et al., *supra* note 142, at 241; Prescott & Spier, *supra* note 136, at 69.

145. Gibson, *supra* note 7, at 945–46; Prescott & Spier, *supra* note 136, at 69.

146. Cooter et al., *supra* note 142, at 225–27.

147. *Id.* at 226, 231–33.

148. Priest & Klein, *supra* note 142, at 16–17.

149. Gibson, *supra* note 7, at 887; *see* Prescott & Spier, *supra* note 136, at 96, 107–08.

holder statutory damages of $150,000 per instance of infringement.[150] For companies that have trained models on millions of copyrighted works, this creates the potential for catastrophic liability. For example, in *Getty Images*, if Stability AI's fair use defense fails, the company could face billions in statutory damages—making settlement a far more appealing and financially sound option.[151]

Aside from exposure to damages, the burdens of litigation make settlement a pragmatic choice.[152] Copyright lawsuits often require years of sustained legal expenses and burdensome discovery, draining valuable resources necessary for product development and operations.[153] Costly procedural burdens alone can create powerful incentives to settle, especially where one party can extract value regardless of legal merits.[154] For companies investing heavily in AI infrastructure and research, settlement offers a faster and more cost-efficient resolution than prolonged litigation, allowing them to avoid disruption while continuing to scale.[155]

Beyond individual exposure, both AI companies and rightsholders are acutely aware of the broader implications of litigation. For dominant AI firms, a single unfavorable court ruling—such as a court finding that generative AI training does not qualify as fair use—could trigger sweeping consequences across the industry, including mandatory licensing regimes and constrained access to training data.[156] Even companies confident in their defenses may settle to avoid establishing a precedent that could threaten long-term innovation.[157] Conversely, rightsholders also have strong incentives to avoid court decisions that might enshrine expansive fair use interpretations. They pursue private settlements that impose favorable licensing terms while preserving control over their work.[158] These agreements allow copyright holders to impose favorable

150. 17 U.S.C. § 504(c)(2) (2022).

151. *See generally* Getty Images Demand, *supra* note 12, ¶¶ 34–36 (describing plaintiff's allegations that defendant Stability AI copied millions of copyrighted images, associated captions, and metadata without authorization and used them to train its AI models).

152. *See* Anup Malani & Jonathan S. Masur, *Raising the Stakes in Patent Cases*, 101 GEO. L.J. 637, 682–83 (2013); *see generally* Gibson, *supra* note 7, at 887, 940–42 (discussing the burdens of litigation); Prescott & Spier, *supra* note 136, at 137–38 (considering the benefits of settlement agreements in civil litigation).

153. Soni, *supra* note 4; *Understanding the Cost of Copyright Infringement Lawsuits*, IPISC (June 28, 2024), https://ipisc.com/understanding-the-cost-of-copyright-infringement-lawsuits.

154. Malani & Masur, *supra* note 152, at 656–57; Prescott & Spier, *supra* note 136, at 78.

155. Soni, *supra* note 4.

156. *See* Malani & Masur, *supra* note 152, at 641; *see generally* CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 3 (2025); Timothy B. Lee, *Copyright Lawsuits Pose a Serious Threat to Generative AI*, UNDERSTANDING AI (Mar. 30, 2023), https://www.understandingai.org/p/copyright-lawsuits-pose-a-serious; *Researchers Say AI Copyright Cases Could Have Negative Impact on Academic Research*, GA. TECH. (Nov. 21, 2024), https://research.gatech.edu/researchers-say-ai-copyright-cases-could-have-negative-impact-academic-research.

157. Gibson, *supra* note 7, at 941–42 ("Combine these doctrinal gray areas and severe consequences with the risk aversion that pervades key copyright industries, and the result is a practice of securing copyright licenses even when none is needed. Better safe than sued.").

158. *Id.* at 884, 899.

licensing terms and shape emerging norms around data use without subjecting their claims to judicial scrutiny or invalidation.[159]

Maintaining licensing relationships with rightsholders ensures continued access to valuable training data for AI companies.[160] The Associated Press and OpenAI licensing deal exemplifies this approach, allowing OpenAI to train on news archives while ensuring compensation for journalists.[161] Moreover, some copyright holders, especially in the publishing industry, prefer settlements that lead to structured licensing deals rather than blanket restrictions on AI training.[162] These agreements enable the parties to resolve disputes cooperatively while preserving commercial relationships and reducing legal uncertainty.

Reputational harm is another powerful driver of settlement. AI companies that pursue aggressive litigation against copyright holders risk public backlash and negative media coverage.[163] OpenAI's high-profile dispute with The New York Times has generated widespread criticism and sparked broader concerns about AI's impact on journalism and creative industries.[164] Because many AI firms depend on public trust to attract users, business partners, and investors, prolonged legal battles—particularly with artists, musicians, and journalists—can damage their brand and erode public trust.[165] As a result, settling disputes privately often provides a more attractive and strategic path for preserving reputation and maintaining long-term goodwill.[166]

2. Outliers or Strategic Holdouts—The Companies That Choose to Fight

Despite the trend toward settlements, some AI companies have chosen to litigate rather than negotiate. These firms may take this approach for several reasons. Some companies, particularly those with extensive legal resources, may be confident in their fair use defense and seek a court ruling that could provide a favorable precedent. Even with recent case law, such as the *Thomson Reuters*

159. Depoorter, *supra* note 13, at 960, 976; Gibson, *supra* note 7, at 884.

160. *Associated Press*, REUTERS, *supra* note 15; Roth, *supra* note 15.

161. *Id.*

162. *See, e.g.*, Miller & Bass, *supra* note 96 (noting that academic publishers are increasingly entering AI licensing agreements to monetize their content and create new revenue streams); *Uncharted Territory: AI and the Cambridge Approach for Academic Book Publishing*, *supra* note 97 (explaining that publishers are implementing opt-in licensing frameworks that allow authors to authorize AI uses of their works in exchange for compensation, reflecting a preference for structured licensing over outright exclusion).

163. *See* News/Media Alliance, *supra* note 89; Torey Akers, *Artists Amoako Boafo, Hans Haacke and Deborah Butterfield Among Thousands to Sign Statement Against AI Content Scraping*, ART NEWSPAPER (Oct. 24, 2024), https://www.theartnewspaper.com/2024/10/24/artists-statement-opposing-artificial-intelligence-content-scraping.

164. *See, e.g.*, Bobby Allyn, *Judge Allows 'New York Times' Copyright Case Against OpenAI to Go Forward*, NPR (Mar. 26, 2025, at 18:28 ET), https://www.npr.org/2025/03/26/nx-s1-5288157/new-york-times-openai-copyright-case-goes-forward (illustrating how high-profile litigation has amplified public scrutiny of AI companies and heightened concerns about harm to journalism and creative sectors).

165. *Id.*

166. Depoorter, *supra* note 13, at 960, 976; Gibson, *supra* note 7, at 884.

decision casting doubt on the viability of fair use in specific AI contexts, enterprises may still view litigation as a strategic opportunity to shape doctrine in developing generative AI models. A ruling favoring generative AI training under fair use would benefit all AI companies, reducing the risk of future lawsuits.

AI firms may also resist settlement to avoid restrictive licensing precedents. A major concern is that privately resolving copyright disputes reinforces the expectation that AI training must be licensed, creating a de facto licensing regime that forces developers to pay rightsholders for access to training data. In areas of legal uncertainty, especially in emerging technologies, settlement agreements can function as vehicles for private parties to shape legal norms while bypassing the role of courts.[167]

Additionally, some AI defendants continue to litigate to challenge expansive copyright claims. Industries such as music and stock photography have historically relied on aggressive enforcement strategies, making them more likely to push for restrictive settlement terms.[168] By resisting settlements, AI firms seek to limit the expansion of copyright protections that could stifle AI development.[169]

The decision to settle or litigate reflects a broader tension between AI innovation and copyright enforcement. While many AI firms choose settlements to mitigate risk, avoid precedent-setting rulings, and maintain business relationships, others resist in an effort to establish legal clarity. However, bargaining in the shadow of the law allows powerful parties to shape legal norms through settlement agreements rather than judicial rulings.[170] The cumulative effect of these settlements is the gradual construction of a "pay-to-train" licensing model that AI companies must navigate.[171] This increases the risk that private agreements will become de facto regulatory standards rather than legislative or judicially defined copyright principles.

B. HOW COPYRIGHT HOLDERS USE SETTLEMENTS TO ASSERT CONTROL

The rise of generative AI has forced copyright holders to confront how their works are used to train models capable of producing text, images, and music. Although some rightsholders have pursued litigation, many prefer private settlements as a means of steering generative AI development without the risks associated with judicial rulings.[172] These agreements allow them to establish

167. *See* Depoorter, *supra* note 13, at 974–76; Gibson, *supra* note 7, at 887.

168. *See, e.g.*, Getty Images Demand, *supra* note 12, ¶¶ 1–2; Concord Music Group Complaint & Demand, *supra* note 15, ¶¶ 1–7.

169. Microsoft Corp., Comment Letter on Notice of Inquiry and Request for Comment re Artificial Intelligence and Copyright (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-8750.

170. Prescott & Spier, *supra* note 136, at 127.

171. *See* Xuan-Thao Nguyen & Elizabeth Porter, *The AI Input Class: Constitutional Urgency and Fair Licencing in AI Copyright Class Actions*, 14 N.Y.U. J. INTELL. PROP. & ENT. L. 1, 26, 85 (2025).

172. Abbott & Rothman, *supra* note 142, at 1166; Bannigan et al., *supra* note 25, at 2.

licensing terms, impose content restrictions, and maintain influence over how generative AI companies innovate.[173] In doing so, rightsholders go beyond merely protecting their existing rights, actively shaping conditions under which AI technologies can develop. These interventions reflect legal, economic, and strategic motivations, often tailored to the distinct risks posed by generative AI.

1. Monetization, Output Restrictions, and Licensing Leverage

A key motivation for rightsholders in pursuing settlements is the ability to convert AI's reliance on copyrighted material into a sustainable revenue stream.[174] The publishing industry, for instance, has framed AI training as a business activity that should require a license, thereby positioning AI companies as paying business partners.[175] The Associated Press's recent licensing agreement with OpenAI, granting access to AP's archives, illustrates this approach.[176] Such deals enable rightsholders to monetize existing content while retaining significant control over its use.[177]

Control over content is equally important, particularly in sectors where generative AI poses a risk of substituting for original works.[178] In the visual arts and music industries, some major rightsholders have taken aggressive stances against AI companies, concerned that AI-generated images and lyrics could saturate the market and erode the value of human-created work.[179] Getty Images, for example, has not only sued Stability AI but has also pursued licensing agreements designed to monetize its extensive database or to prevent unauthorized AI usage altogether.[180] Similarly, major music publishers have focused on securing agreements that prevent models from generating outputs resembling copyrighted songs.[181]

Avoiding adverse judicial rulings is another major driver behind rightsholders' preference for settlements. A court decision affirming that generative AI training qualifies as fair use would severely weaken rightsholders'

173. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 60 (2025).

174. Pasquale & Sun, *supra* note 17, at 230.

175. *See* U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 34 n.199 (2025) (quoting Associated Press, Comment Letter on Artificial Intelligence and Copyright Study (Oct. 27, 2023), https://www.regulations.gov/comment/COLC-2023-0006-8360) ("Because one purpose of news publishers' content is to license it to AI developers for model training, the use by AI developers 'share[s] the objectives' of news publisher's content.").

176. O'Brien, *supra* note 86.

177. *See* Brittain, *supra* note 15; *see also* Stipulation and Order Regarding Preliminary Injunction Motion ¶ 5–7, Concord Music Grp., Inc. v. Anthropic PBC, No. 5:24-CV-03811 (N.D. Cal. Jan. 2, 2025) (No. 291).

178. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 65 (2025); *see also* Bannigan et al., *supra* note 25, at 4.

179. *See* U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 65 (2025); Getty Images Demand, *supra* note 12, ¶ 36; Concord Music Group Complaint & Demand, *supra* note 15, ¶¶ 6, 66, 69.

180. Press Release, Getty Images, Getty Images Statement (Jan. 17, 2023), https://newsroom.gettyimages.com/en/getty-images/getty-images-statement.

181. *See* Brittain, *supra* note 15.

leverage in negotiating licensing agreements. Though the recent decision in *Thomson Reuters* rejected a fair use defense—bolstering rightsholders' confidence in litigation—the ruling was limited to a non-generative context, leaving open how courts will assess training for generative models.[182] As a result, many rightsholders prefer the predictability of private negotiations, which allows them to impose licensing terms without risking a court decision that could narrow their control over future disputes.

Rightsholders are further motivated by concerns over market disruption—specifically, generative AI's impact on traditional revenue streams.[183] News organizations warn that AI-generated summaries trained on their reporting could threaten subscription-based models, reducing readership and advertising revenue as a result.[184] Similarly, book publishers fear AI-generated narratives could undercut demand for traditional literature.[185] These threats have prompted many rightsholders to push for settlements imposing AI-generated output restrictions or mandating financial compensation through licensing fees.[186]

Although strategies differ across sectors, the underlying goals align. News and publishing firms generally seek licensing revenue and attribution.[187] Visual arts organizations focus on limiting generative AI model capabilities to prevent direct replication and competition with human works.[188] Music publishers have prioritized output restrictions over monetary compensation to protect their markets.[189] Each industry's approach reflects unique vulnerabilities posed by the rapidly evolving ecosphere of generative AI.

2. Quiet Rulemaking Without Courts or Congress

Through settlements, rightsholders effectively dictate the legal terms under which generative AI develops. Large corporations like The New York Times, Getty Images, and Universal Music Group have sufficient resources to secure favorable terms, whereas independent creators often lack comparable bargaining power or means to influence outcomes, secure protections, or even understand how their works are used.[190] This disparity raises substantial concerns regarding

182. Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc., 765 F. Supp. 3d 382, 399–401 (D. Del. 2024).

183. *See* U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 64–71 (2025).

184. *See supra* notes 79, 85.

185. *See* Kadrey v. Meta Platforms, Inc., 788 F. Supp. 3d 1026, 1051 (N.D. Cal. 2025).

186. *See* discussion *supra* Subpart.I.B.

187. *See* discussion *supra* Subparts.I.B.1, I.B.2.

188. *See* discussion *supra* Subpart.I.B.3.

189. *See* discussion *supra* Subpart.I.B.4.

190. *See generally* Pasquale & Sun, *supra* note 17, at 230 (discussing how the withdrawal of major institutional rightsholders like The New York Times would significantly impact training data quality, underscoring their leverage in negotiations over AI training use); Jorge Padilla & Kadambari Prasad, *Demystifying Licensing Debtates: Should GenAI Developers Pay to Train Their Models on Copyright Protected Content?*, COMPASS LEXECON, Feb. 25, 2025, at 5–8, https://compass-lexecon.files.svdcdn.com/production/editorial/2025/02/The-Analysis-GenAI-250225.pdf?dm=1740477422 (explaining that licensing terms are negotiated bilaterally and vary across rightsholders, reflecting differences in bargaining power).

whether the emerging regulatory regime disproportionately favors large rightsholders at the expense of independent creators, driven more by market power than principles of fairness or transparency.

These private agreements not only reflect bargaining disparities but also shape broader industry expectations. By resolving disputes confidentially, major rightsholders craft informal licensing norms that AI developers feel compelled to follow to avoid litigation, especially absent definitive judicial rulings addressing generative AI.[191] The growing reliance on settlements reflects a marked shift away from asserting fair use arguments, instead reinforcing the view that generative AI training must require a license.[192]

This movement toward mandatory licensing carries significant legal and economic implications. If licensing becomes the default, fair use arguments risk becoming untenable for AI developers.[193] Such a development would effectively create a pay-to-train system where only well-funded AI companies can afford to develop sophisticated models, reducing competition and innovation in the AI space.[194] Allowing generative AI copyright norms to develop primarily through private agreements rather than judicial or legislative intervention risks producing a regulatory framework driven more by market leverage than legal clarity.[195]

Beyond structural and market implications, the privatized nature of these settlements raises profound questions about democratic legitimacy. When legal standards are established through confidential agreements among a narrow set of corporate actors, the public is effectively excluded from participating in or even observing critical rulemaking processes.[196] The resulting norms, though lacking formal statutory authority, govern generative AI development with comparable practical authority. This displacement of judicial and legislative oversight deprives the public of procedural safeguards, interpretive transparency, and democratic accountability.[197] In effect, copyright law is quietly redefined in corporate boardrooms rather than openly in courtrooms or legislatures, thereby calling the legitimacy of this emerging legal regime into question.

While settlements offer financial certainty for rightsholders and legal predictability for major AI enterprises, they simultaneously limit judicial

191. CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 3 (2025).

192. Nguyen & Porter, *supra* note 171, at 26.

193. Tori Noble, *AI and Copyright: Expanding Copyright Hurts Everyone—Here's What to Do Instead*, ELEC. FRONTIER FOUND. (Feb. 19, 2025), https://www.eff.org/deeplinks/2025/02/ai-and-copyright-expanding-copyright-hurts-everyone-heres-what-do-instead.

194. *See id.*

195. *See* Nguyen & Porter, *supra* note 171, at 26.

196. *See id.* at 57.

197. FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 216–17 (2015) (discussing how pervasive secrecy in important decision-making can undermine public values); *see* Mark MacCarthy, *Copyright Alone Cannot Protect the Future of Creative Work*, BROOKINGS (May 1, 2025), https://www.brookings.edu/articles/copyright-alone-cannot-protect-the-future-of-creative-work.

interpretation and broader public discourse about fairness and accessibility within the AI copyright domain. As these private agreements continue to influence industry standards, the legal status of training generative AI models remains unresolved, underscoring the persistent uncertainty over whether such practices qualify as fair use or require formal licensing.[198]

C. A SHADOW LEGAL SYSTEM AND ITS STRUCTURAL CONSEQUENCES

As generative AI copyright disputes increasingly resolve through confidential settlements, these private agreements become substitutes for formal law. Rather than merely ending specific conflicts, they establish informal norms governing training data access, output behavior, and licensing expectations. Though privately negotiated, these norms now influence the design and deployment of AI systems across the industry.[199]

This phenomenon has led to the creation of a quasi-regulatory regime—a system where copyright holders and prominent AI developers dictate compliance terms without legislative or judicial clarity.[200] Licensing deals with content-rich platforms like Shutterstock and the Associated Press illustrate how AI companies are adapting their conduct based on privately imposed conditions around data access and compensation structures.[201]

This privatized rulemaking displaces not only courts and legislatures but also raises serious competition concerns. The emerging settlement-driven regime disproportionately benefits incumbents with sufficient capital and legal resources to absorb licensing costs or negotiate favorable terms.[202] In contrast, smaller AI firms face prohibitive barriers to entry not because of technical inferiority but because of private arrangements that gatekeep essential training materials.[203] As a result, private copyright enforcement reshapes the competitive terrain of AI, transforming access to training data into a function of financial leverage rather than technological merit.[204] This exclusionary dynamic undermines copyright's public justification of promoting innovation and instead

198. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 105 (2025); *see* Brittain, *supra* note 15.

199. Depoorter, *supra* note 13, 958–59.

200. *See* Pamela Samuelson, *Legally Speaking: The Dead Souls of the Google Book Search Settlement*, 52 COMMC'NS. ACM 28, 28 (2009) (describing the Google Book Search settlement as a privately negotiated compulsory license that restructures access to copyrighted works outside the formal legislative process).

201. Roth, *supra* note 15; *Associated Press*, REUTERS, *supra* note 15.

202. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 34 (2025).

203. *Id.* at 74, 74 n.417, 75 (quoting Andreessen Horowitz, Comment Letter on Notice on Inquiry on Artificial Intelligence & Copyright (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-9057) ("[T]he cost of paying to license even a fraction of the content needed to properly train an AI model would be prohibitive for all but the deepest-pocketed AI developers, resulting in dominance by a few technology incumbents. This would undermine competition by the technology startups which are the source of the greatest innovation in AI.").

204. *See id.*; Soni, *supra* note 4.

reinforces market power among a handful of dominant actors.[205] Scholars observe this creates a "selection effect" whereby repeat players systematically consolidate their advantages over smaller, risk-averse competitors.[206]

These settlements do not merely resolve liability; they communicate industry-wide expectations that others feel compelled to follow to avoid litigation.[207] Because parties negotiate concerning their expected legal entitlements, these agreements often replace court rulings by embedding projected trial outcomes into private standards that carry industry-wide force.[208]

Several of these agreements explicitly impose restrictions on AI model outputs, underscoring their function as private rulemaking tools. For example, the deal between Anthropic and music publishers—imposing technical guardrails that prohibit the generation of song lyrics—effectively dictates product features through the lens of copyright risk.[209] In this manner, terms formulated during private negotiations become embedded directly into model architectures and training protocols.[210] This shift means key decisions about what AI systems can learn, produce, or disseminate are increasingly made behind closed doors—not in public legal forums. This concentration of normative power in the hands of a small group of private actors, arising through confidential bargaining rather than transparent legal processes, raises important questions about democratic legitimacy and public accountability.

The evolving landscape of private settlements shapes not only technical development but also the broader legal environment.[211] As AI developers consistently opt for settlement over litigation, courts lack opportunities to clarify whether training generative models constitutes fair use. Continued reliance on private settlements perpetuates underlying legal ambiguity that favors rightsholders in negotiation, leading to a gradual expansion of copyright power by risk-averse behavior and excessive licensing.[212] This dynamic is amplified when well-resourced actors can internalize litigation costs and exploit uncertainty to secure advantageous terms, further entrenching bargaining asymmetries in future disputes.[213]

---

205. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 74–75 (2025); Mantegna, *supra* note 94, at 1167.

206. Marc Galanter, *Why the "Haves" Come Out Ahead: Speculations on the Limits of Legal Change*, 9 L. & SOC'Y. REV. 95, 97–104 (1974); Timothy Meyer, *Codifying Custom*, 160 U. PA. L. REV. 995, 1050–51 (2012).

207. Depoorter, *supra* note 13, at 958–59.

208. Cooter et al., *supra* note 142, at 227–29; Samuelson, *supra* note 200.

209. *See* Brittain, *supra* note 15.

210. *See, e.g.*, Pamela Samuelson, *Thinking About Possible Remedies in the Generative AI Copyright Cases*, COMMC'N. ACM (forthcoming 2024) (manuscript at 6), https://ssrn.com/abstract=4770671; Shani Rivaux, Macarena Ferreira Fink & Marcus Leonard, *Lyric or Leave It: Anthropic Tries to Strike a Chord with the Music Industry*, PILLSBURY (Feb. 4, 2025), https://www.pillsburylaw.com/en/news-and-insights/anthropic-copyright-claude-ai.html; Brittain, *supra* note 15.

211. *See* Cooter et al., *supra* note 142, at 247; Malani & Masur, *supra* note 152, at 641.

212. *See* Gibson, *supra* note 7, at 898–900. Cooter et al., *supra* note 142, at 247.

213. *See* Cooter et al., *supra* note 142, at 241; *see also* Prescott & Spier, *supra* note 136, at 69.

These imbalances intensify when repeat-player institutions face single-case defendants whose heightened risk aversion makes them likely to concede, even when a viable defense exists.[214] Yet this ambiguity is not simply a product of judicial delay; it is sustained by a selection effect, where cases least likely to produce a clear legal resolution are the ones most likely to be litigated, while more straightforward disputes are quietly settled.[215] As AI companies more commonly accept licensing as necessary, even when not legally mandatory, this expectation becomes more firmly anchored within the market.[216]

The widespread adoption of licensing agreements exerts cascading effects on market structure and competition. Licensing settlements are typically negotiated by large, well-resourced companies—such as OpenAI, Microsoft, and Google—that can afford multimillion-dollar deals.[217] Startups and smaller firms, by contrast, lack the leverage or capital to participate in these negotiations, effectively excluding them from the training data needed to develop competitive models.[218] Dominant firms exploit this uncertainty to solidify monopolistic advantages in ways that raise costs for competitors in the space.[219] The resulting pay-to-train environment thus creates substantial structural barriers to entry, directly conflicting with copyright's foundational purpose of fostering innovation.

This chilling effect impacts more than just market entry but also stifles innovation itself. Companies may hesitate to experiment with AI systems that could qualify for fair use protections due to fear of potential litigation. Without judicial guidance, firms often resort to self-regulatory practices that prioritize compliance over innovation.[220] This dilemma disproportionately harms research institutions, nonprofit developers, and small teams who lack the resources for extensive legal reviews or to secure tailored licenses for every dataset they consider using.[221]

---

214. *See* Meyer, *supra* note 206, at 1050–51; *see also* Malani & Masur, *supra* note 152, at 640.

215. Priest & Klein, *supra* note 142.

216. *See* Gibson, *supra* note 7, at 899.

217. Bertin Martens, *Economic Arguments in Favour of Reducing Copyright Protection for Generative AI Inputs and Outputs* 6 (Bruegel, Working Paper No. 09/2024), https://www.bruegel.org/system/files/2024-04/WP%2009%20040424%20Copyright%20final_0.pdf; *see* Giuseppe Colangelo, *A Competition Policy Analysis of Copyright Protection in Generative AI*, SING. J. LEGAL STUD., Sep. 2025, at 7–8.

218. Martens, *supra* note 217; *see* Quang, *supra* note 40, at 1425–26 (discussing resource disparities between established AI firms and smaller innovators).

219. *See* U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 74–75 (2025); Martens, *supra* note 217.

220. Gibson, *supra* note 7, at 887; Michal Shur-Ofry, *Hatch-Waxmanizing Copyright*, 18 MICH. TELECOMMS. & TECH. L. REV. 171, 177–78 (2011).

221. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 75 n.419 (2025) ("While mid to large-size businesses have the financial means and workforce to hire legal teams to track down and obtain copyright use permissions . . . [e]ntrepreneurs, start-ups, and small businesses do not have the financial means or workforce to obtain permission from copyright owners." (quoting Regulosity & Pangea, Comment Letter on Artificial Intelligence and Copyright (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-9039)).

Left unchallenged, this settlement-driven process may also entrench outdated understandings of copyright law. The unique attributes of generative AI, such as the lack of human authorship and the capacity to produce novel content, challenge traditional copyright frameworks.[222] If these compliance issues remain settled privately rather than adjudicated, critical doctrinal questions may never reach courts. As a result, legacy assumptions about authorship, originality, and infringement persist without scrutiny, even as generative AI transforms the way content is created and consumed.[223]

Thus, the rise of private settlements in generative AI copyright disputes has produced a shadow framework that governs how AI systems are trained and what they can output. This informal framework operates with the force of law without the procedural safeguards or public accountability of the judicial or legislative process.[224] Its long-term effects—market consolidation, reduced access to information, legal stagnation, and weakened fair use protections—raise pressing questions about whether the future of AI copyright should be dictated by private leverage or public law.[225] Over time, avoiding judicial decisions risks causing courts and regulators to treat privately negotiated licensing norms as de facto legal consensus, thereby retrenching expansive copyright interpretations where doctrinal evolution is necessary.

III. POLICY OPTIONS AND THEIR IMPLICATIONS FOR THE FUTURE

A. CODIFYING FAIR USE TO DISRUPT THE PAY-TO-TRAIN REGIME

One path to dismantling the emerging settlement-driven system is for Congress to codify fair use for AI training, explicitly legalizing the use of copyrighted works to develop generative AI models. Codification would eliminate the core leverage rightsholders have relied on to pressure AI enterprises into licensing agreements, ultimately preventing private settlements from continuing to define AI copyright law.

Currently, AI firms settle because of legal uncertainty.[226] Without clear judicial precedent affirming that AI training qualifies as fair use, rightsholders have been able to assert copyright infringement claims and compel AI companies to adopt licensing agreements on their terms.[227] Legal scholars have shown that ambiguity disproportionately benefits rightsholders in settlement negotiations, enabling them to extract licensing fees even when their claims

222. *See* CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 1–2 (2025); Hutson, *supra* note 21, at 896.

223. *See, e.g.*, Thaler v. Perlmutter, 687 F. Supp. 3d 140, 146–47 (D.D.C. 2023) (reaffirming human authorship as copyright requirement); Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc., 765 F. Supp. 3d 382, 391 (D.Del. 2024).

224. Owen M. Fiss, *Against Settlement*, 93 YALE L.J. 1073, 1075–78 (1984).

225. *See* U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 74–75 (2025); Shur-Ofry, *supra* note 220, at 173.

226. *See supra* Subpart.II.A.1.

227. *See supra* Subpart.I.B.

might not withstand judicial scrutiny.[228] If Congress were to legislate that AI training falls within fair use, this bargaining imbalance would disappear, and AI companies would no longer need to bargain in the shadow of the law for the right to train their models.[229]

The foundation for such a reform exists in previous judicial interpretations of fair use, particularly in cases where courts have recognized that transformative uses of copyrighted works, such as *The Authors Guild v. Google*, can be lawful.[230] Scholars have argued that generative AI training serves a comparable function: Rather than directly competing with original works, AI models analyze vast datasets to generate novel outputs, much like how search engines utilize copyrighted material to yield relevant results without infringing.[231]

Beyond eliminating unnecessary licensing agreements, codifying AI training as fair use would challenge settled notions of authorship and ownership in AI-generated content.[232] While rightsholders contend that their works are being used to generate new content, training data typically serves to inform model behavior rather than directly reproduce protected expression, weakening the case for mandatory licensing.[233] If training data merely informs the generation of novel, non-infringing outputs, there is little basis to treat such use as derivative works.[234] Codifying fair use could clarify that training on copyrighted works—absent substantial reproduction of such works—does not inherently require rightsholder approval, ensuring that AI companies are not automatically forced into licensing schemes.[235]

By codifying AI training as fair use, Congress would disrupt the licensing framework that private settlements have established, reducing the leverage rightsholders currently use to push AI companies into pay-to-train agreements. This shift could lower compliance costs and reopen the AI development landscape to smaller firms, allowing them to compete without the financial burdens that currently favor well-funded industry leaders like OpenAI and Google.[236]

---

228. Gibson, *supra* note 7, at 945.

229. *See* Gibson, *supra* note 7, at 945.

230. *See, e.g.*, Authors Guild v. Google, Inc., 804 F.3d 202, 207–08 (2d Cir. 2015).

231. *See* Lee, *supra* note 25, at 1457–58; Mantegna, *supra* note 94, at 1162; *see also* CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 1–2 (2025).

232. Lee, *supra* note 25, at 1493; Mantegna, *supra* note 94, at 1163.

233. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 6 (2025); Mantegna, *supra* note 94, at 1165; Lee, *supra* note 25, at 1580.

234. *See* U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 20 (2025); *see also* Mantegna, *supra* note 94, at 1147.

235. *See* Joshua Levine, *To Preserve US AI Leadership, Congress Must Address Copyright*, TECH POL'Y PRESS (Mar. 13, 2025), https://www.techpolicy.press/to-preserve-us-ai-leadership-congress-must-address-copyright.

236. *See* Depoorter, *supra* note 13, at 980–84 (discussing how legal uncertainty enables rightsholders to extract settlements).

However, such reform would likely face strong resistance. Sectors that have embraced licensing as a default compliance model—particularly the publishing and visual arts industries—will likely oppose a statute that permits training without compensation. While some judicial interpretations of fair use support the legality of data use for transformative purposes, courts have not yet ruled definitively on whether generative AI training qualifies.[237] If Congress enacts fair use protections prematurely, rightsholders may challenge the law in court, leading to prolonged litigation that could weaken the impact of the reform.[238]

Even with legal protection for AI training, risks would remain for generative models. If AI models produce content that closely mimics copyrighted works, rightsholders may still bring infringement claims based on their outputs rather than the training inputs.[239] As a result, codifying fair use for training would not eliminate the threat of litigation entirely.

The proposed reform would also have ripple effects. Reducing reliance on settlements could encourage courts to engage more directly with unresolved fair use questions, reintroducing public legal reasoning into an area now dominated by private agreements. At the same time, this shift could destabilize existing licensing arrangements, prompting pushback from stakeholders who have already secured compensation through settlement.[240] Greater judicial engagement may also lead to unpredictable outcomes, increasing uncertainty for rightsholders and AI developers alike.

Ultimately, codification of fair use would represent the most interventionist policy route—one that would replace the current system with a framework where disputes are resolved publicly through the courts and guided by explicit legislation. Whether Congress is prepared to enact such a measure remains uncertain. In the absence of legislative action, however, the de facto licensing regime enforced through settlements will likely continue to expand, making it increasingly difficult for AI firms to operate outside private agreements with rightsholders.[241]

B. TRANSPARENCY AND OPT-OUT MECHANISMS AS A MIDDLE GROUND

While codifying fair use would provide AI firms with broad legal protection, such a reform faces significant political and legal challenges as rightsholders are likely to oppose it fiercely. Alternatively, a transparency and opt-out framework offers a middle-ground solution that balances the interests of both AI developers and copyright holders. Inspired by the European Union

237. Gibson, *supra* note 7, at 888–89.

238. *See* Lee, *supra* note 25, at 1574.

239. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 29–30 (2025).

240. *See* Samuelson, *supra* note 210.

241. *See* Depoorter, *supra* note 13, at 974.

Copyright Directive, this approach would require AI companies to disclose their training data sources.[242] A transparency regime could mitigate the unchecked power of privately negotiated licensing arrangements while preventing an outright ban on AI training without licensing, since rightsholders would need to assert their rights proactively rather than impose blanket restrictions.[243] Implementing this transparency and opt-out approach would help correct the current imbalance, where settlements effectively serve as informal copyright policies adopted without judicial or legislative oversight.[244]

AI firms currently operate under opaque data collection practices, often refusing to disclose the sources of their training datasets.[245] This lack of transparency allows rightsholders to make sweeping infringement claims, which in turn pressures companies into settlements—even when the alleged use may not clearly implicate protected expression.[246] Mandating disclosure of training data would shift the dynamics of AI copyright enforcement. Without transparency, rightsholders must rely on speculative litigation to assert control over their works.[247] If developers were required to reveal the datasets used to train their models, rightsholders could more precisely evaluate whether their works were used and whether that use potentially constitutes infringement.[248]

---

242. *See generally* Council Regulation 2024/1689, art. 53(1)(d), 2024 O.J. (L 144) 28; Advant-Nctm, *The European Commission's Template on Training Data Transparency: First Guidelines for the AI Act*, LEXOLOGY (Apr. 17, 2025), https://www.lexology.com/library/detail.aspx?g=7d91293e-704b-48f5-92d0-c21ca61b0e26 ("The regulation requires providers of such models to make publicly available a *sufficiently detailed summary* of the data used for training . . . .").

243. Jennifer E. Rothman, *Copyright's Private Ordering and the "Next Great Copyright Act"*, 29 BERKELEY TECH. L.J. 1595, 1597–98 (2014); *see* DEV GANGJEE ET AL., OXFORD INTELL. PROP. RSCH. CTR., OPEN CONSULTATION ON COPYRIGHT AND ARTIFICIAL INTELLIGENCE 4–5 (2025).

244. *See* Ben Depoorter, *Technology and Uncertainty: The Shaping Effect on Copyright Law*, 157 U. PA. L. REV. 1831, 1860 (2009) ("[F]or representatives of new technologies, it can be useful to prolong periods of uncertainty, as those periods leave more time for users of new technology to begin to incorporate the new technology into their everyday lives and to entrench the norm of free content in the public's perception."); Rothman, *supra* note 243; Taylor Bussey, "*You Got Too Much Dip on Your Chip!" How Stagnant Copyright Law Is Stifling Creativity*, 27 J. INTELL. PROP. L. 277, 279 (2020).

245. Zuzanna Warso & Maximilian Gahntz, *How the EU AI Act Can Increase Transparency Around AI Training Data*, TECH. POL'Y PRESS (Dec. 9, 2024), https://www.techpolicy.press/how-the-eu-ai-act-can-increase-transparency-around-ai-training-data; Ana Andrijevic, *Generative AI and Transparency of Databases and Their Content, From a Copyright Perspective*, CTR. ART L. (May 21, 2024), https://itsartlaw.org/2024/05/21/generative-ai-and-transparency-of-databases-and-their-content-from-a-copyright-perspective/#:~:text=This%20contribution%20focuses%20on%20the,the%20principle%20of%20transparency%20angle.

246. *See* Ben Depoorter, *Copyright Enforcement in the Digital Age: When the Remedy is the Wrong*, 66 UCLA L. REV. 400, 404–06 (2019) (showing how "the availability of statutory damages tempts plaintiffs into asserting dubious infringement claims . . . " and pushing for settlements); Bussey, *supra* note 244.

247. *See generally* U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 29 (2025) (noting that courts addressing AI training claims have "reached varying conclusions," reflecting legal uncertainty driving reliance on litigation); *see, e.g.*, Advant-Nctm, *supra* note 242 (explaining that uncertainty around AI training data has led rightsholders to pursue litigation to assert control over potential uses of their works).

248. *Id.*; Anthony Leung, *To Opt Out or Not to Opt Out—The Question of the "Opt-Out Model" for AI Training*, MONDAQ (April 21, 2025), https://www.mondaq.com/article/1610498.

Mandating this disclosure would narrow the scope of legal uncertainty and allow both sides to calibrate their legal positions more accurately.

Greater transparency would also discourage rightsholders from making overbroad claims. AI firms could demonstrate that their models were trained on public domain content, licensed datasets, or other materials that do not implicate copyright protection.[249] Increasing access to such factual information weakens the negotiating power of rightsholders, who often benefit from ambiguity in settlement negotiations.[250] By introducing a formal disclosure requirement, a transparency framework could limit speculative claims and reduce unnecessary licensing demands.[251]

However, transparency proposals face opposition from industry leaders. AI firms argue that their training datasets contain proprietary business information and that disclosure could undermine their competitive advantage by revealing trade secrets or competitive strategies.[252] As such, some companies, including OpenAI, have declined to disclose their training sources.[253] Regulatory measures could nonetheless be structured to safeguard sensitive business information while ensuring a basic level of accountability in developing generative AI models.

Transparency alone is insufficient to address rightsholders' concerns comprehensively. An opt-out mechanism—designed to complement transparency—could further rebalance control between rightsholders and developers. Modeled after the EU's "text and data mining" directive, such a system would allow copyright holders to exclude their works from AI training datasets proactively.[254] Under this system, rightsholders would register their works in an AI exclusion database, which AI firms are required to consult before training their models.[255] Implementing this system would offer a structured and scalable alternative to the current settlement-driven system, where rightsholders must rely on litigation to negotiate licensing terms.[256]

---

249. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 72 (2025); Katherine Klosek & Marjory S. Blumenthal, *Training Generative AI Models on Copyrighted Works Is Fair Use*, ASS'N OF RSCH. LIBS. (Jan. 23, 2024, at 16:48 ET), https://www.arl.org/blog/training-generative-ai-models-on-copyrighted-works-is-fair-use.

250. Gibson, *supra* note 7, at 889.

251. *See* Buick, *supra* note 22, at 185 (arguing that such transparency will enable rightsholders to enforce their rights in a targeted manner, obviating the need for broad, speculative claims or blanket licensing); *see generally* Generative AI Copyright Disclosure Act of 2024, H.R. 7913, 118th Cong. § 2(1)(a)–(b) (2024) (proposing a disclosure regime requiring identification of copyrighted works used in AI training).

252. *See* Council Regulation 2024/1689, recital 107, 2024 O.J. (L 144) 28 (recognizing the need to protect "trade secrets and confidential business information" in connection with AI training data disclosure); Advant-Nctm, *supra* note 242 (describing stakeholder concerns that disclosure requirements may reveal "strategic assets" such as data collection methods and algorithms).

253. Warso & Gahntz, *supra* note 245; Andrijevic, *supra* note 245.

254. *See* Council Directive 2019/790, art. 4(1)–(3), 2019 O.J. (L 130) 113–14 (establishing a text-and-data-mining exception subject to an opt-out mechanism allowing rightsholders to reserve their rights).

255. Generative AI Copyright Disclosure Act of 2024, H.R. 7913, 118th Cong. § 2(a)(1).

256. *See* Pamela Samuelson, *The Google Book Settlement as Copyright Reform*, 2011 WISC. L. REV. 479, 529 (2011); Rothman, *supra* note 243, at 1597–98.

A well-implemented opt-out model would allow rightsholders to retain control over their works without mandating licensing agreements, ensuring that AI training is not automatically classified as infringement while still granting copyright holders meaningful enforcement mechanisms.[257] Applying such a model could reduce unnecessary lawsuits while preserving the ability of AI developers to access publicly available and non-restricted content for training purposes.[258] Critically, the opt-out framework would challenge the assumption that all AI training requires licensing. Shifting the burden onto rightsholders to opt out balances market power by preventing AI companies from being forced into one-sided licensing agreements to avoid litigation risks.

However, this system presents practical challenges. Unlike in Europe, where the text and data mining exception is codified in law, the United States would need to establish a centralized database for rightsholders to submit their opt-out requests, which could require significant administrative oversight.[259] Moreover, some AI companies may exploit loopholes—arguing that certain types of training data do not fall under the opt-out system—resulting in inconsistent applications and legal uncertainty.[260] These structural hurdles must be addressed to ensure the opt-out model functions as a viable alternative to settlement-driven enforcement.

Despite these challenges, a transparency and opt-out approach offers a viable middle path. If implemented effectively, it could weaken the de facto licensing regime and curtail the unchecked influence of private settlement—all while preserving a competitive development environment for generative AI and respecting the legitimate interests of content creators.

C. LETTING IT RIDE—WHAT HAPPENS IF CONGRESS DOES NOTHING?

If Congress fails to intervene, the rules governing generative AI and copyright will continue to emerge through private negotiations rather than judicial or legislative processes. The growing dominance of confidential settlements and licensing agreements has already laid the groundwork for a de facto regulatory framework where access to training data is dictated by rightsholders, not legal precedent.[261] Without statutory clarity, the pay-to-train model will become further cemented, favoring well-funded AI firms capable of affording licensing deals while marginalizing smaller players.[262]

---

257. *See* Council Regulation 2019/790, recital 2, 2024 O.J. (L 130) 92.

258. *But see* Leung, *supra* note 248 (arguing that adoption of an opt-out model is legally unsound and practically unfeasible because it will likely create more uncertainties).

259. Council Directive 2019/790, art. 4(3), 2019 O.J. (L 130) 114.

260. *See* Leung, *supra* note 248.

261. Lilian Edwards, Igor Szpotakowski, Gabriele Cifrodelli, Joséphine Sangaré & James Steward, *Private Ordering and Generative AI What Can We Learn from Model Terms and Conditions* 4 (CREATe, Working Paper No. 2024/5).

262. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 69–70 n.396 (2025) ("[T]he costs associated with obtaining these licenses could make AI projects excessively expensive, thus impeding innovation and hindering industry growth. This approach may render many AI-driven

This privatized approach to enforcement presents significant consequences. In the absence of public oversight, settlements effectively substitute for formal regulation, influencing industry behavior through licensing norms that lack transparency and accountability. Courts have historically interpreted copyright law to evolve alongside new technologies, particularly in determining the scope of fair use.[263] However, the preference for settlement over litigation has deprived courts of the opportunity to evaluate whether AI training qualifies as fair use, leaving a gap in jurisprudence that only strengthens rightsholders' negotiating power.[264] Over time, this absence of precedent could influence future rulings as courts may begin to view licensing not as a voluntary risk-management tool but as a de facto legal requirement.[265]

This unsettled legal terrain also threatens to fragment the regulatory landscape. In the absence of statutory rules, private settlements vary dramatically across sectors—some prioritizing training data restrictions, others focusing on output limitations or royalty structures.[266] The result is a patchwork system in which AI companies must navigate inconsistent licensing terms, compliance obligations, and litigation threats depending on the content domain at issue. Without harmonized legal guidance, courts in different jurisdictions could issue conflicting rulings, further complicating compliance for developers operating across markets. The firms most equipped to absorb this uncertainty—namely, dominant tech companies—will be best positioned to shape emerging norms.[267]

Perhaps the most enduring consequence of congressional inaction is the risk of deep market consolidation. Licensing costs create barriers to entry for startups and research institutions, many of which cannot afford to license large training datasets or defend themselves in court.[268] As large AI firms increasingly strike exclusive deals with rightsholders, the opportunity to build competitive

projects unattainable, particularly for smaller entities or researchers with limited resources."); *see* Cooter et al., *supra* note 142, at 246–77; Prescott & Spier, *supra* note 136, at 69; Priest & Klein, *supra* note 142; Bertin Martens, *Why Artificial Intelligence Is Creating Fundamental Challenges for Competition Policy*, BRUEGEL (July 18, 2024), https://www.bruegel.org/policy-brief/why-artificial-intelligence-creating-fundamental-challenges-competition-policy?utm.

263. *See* Authors Guild v. Google, 804 F.3d 202, 207 (2d Cir. 2015); Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc., 765 F. Supp. 3d 382, 392–93 (D. Del. 2025).

264. Depoorter, *supra* note 13, at 983; Gibson, *supra* note 7, at 884.

265. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING 34 (2025).

266. *Id.* at 69 n.395 ("Media reports indicate several examples of companies like Reuters and Shutterstock entering into licensing deals with AI developers, but the feasibility of such direct licensing depends on the nature of the works and the concentration of rights in the relevant market. In many instances, transaction costs are likely to be high.").

267. *See* Kristelia A. Garcia, *Private Copyright Reform*, 20 MICH. TELECOMMS. & TECH. L. REV. 1, 7 (2013); Martens, *supra* note 262.

268. Martens, *supra* note 262.

models diminishes for all but the most well-resourced developers.[269] Because smaller firms lack the resources to absorb legal risk, the ambiguity in copyright law gives larger companies a competitive advantage—they can afford to settle, license, or litigate, while others cannot—effectively raising barriers to entry.[270]

In this environment, compliance with copyright law becomes more dependent on financial leverage rather than legal merit. Startups might find themselves compelled to operate in legal ambiguities or to cease training activities while dominant players negotiate conditions that bolster their market supremacy. The consequences do not simply stifle innovation—they reshape the trajectory of generative AI by allowing a handful of corporations to define what is legally permissible and technologically possible.[271] Absent intervention, this licensing-heavy structure will likely become the default, giving private actors free rein to dictate terms that should otherwise be determined through public law.

Therefore, congressional inaction does not necessarily imply a neutral regulatory stance; it merely delegates authority to those with the most market power. As settlements accumulate and licensing increasingly becomes standard practice, legal interpretations of fair use will likely mirror the power dynamics of these negotiations rather than the foundational principles of copyright law. The crucial issue now extends beyond whether Congress should intervene; it concerns how long Congress can afford to delay before private leverage solidifies into enduring legal doctrine. The longer this trend continues, the more challenging it will be to restore judicial or legislative clarity.

CONCLUSION

The rapid rise of generative AI has revealed deep uncertainties in U.S. copyright law, prompting developers and rightsholders to resolve disputes through private settlements rather than judicial or legislative processes. Private agreements have become the dominant mechanism for addressing conflicts over AI training data in the absence of formal legal guidance, gradually reshaping the regulatory environment by entrenching licensing as the default norm. While these deals offer short-term certainty, they effectively displace fair-use protections that have traditionally supported transformative innovation.

These settlements now operate not only as private contracts but as instruments of de facto regulation. Licensing fees, content restrictions, and compliance obligations allow copyright holders to impose terms that disproportionately benefit their interests, often at the expense of independent

269. *See* Lloyd Dixon, Susan M. Gates, Kanika Kapur, Seth A. Seabury & Eric Talley, *The Impact of Regulation and Litigation on Small Business and Entrepreneurship* 1–2 (RAND, Working Paper No. WR-317-ICJ, 2006).

270. *Id.* at 2.

271. Gibson, *supra* note 7, at 884 ("Thus, the practice of licensing within gray areas eventually makes those areas less gray, as the licensing itself becomes the proof that the entitlement covers the use.").

creators and smaller AI developers. This trend risks solidifying a pay-to-train environment in which only well-capitalized firms can afford to compete, thereby stifling innovation and raising barriers to entry.

If courts and lawmakers continue to defer to private ordering, AI copyright law may be shaped entirely by confidential deals rather than transparent legal standards. Such a system would deepen existing power imbalances in both the creative and technology sectors while leaving unresolved the foundational question of whether generative AI training qualifies as fair use.

The future of AI copyright law hinges on whether policymakers intervene to recalibrate this system. Without action—whether through codifying fair use for AI training, enacting transparency and opt-out mechanisms, or clarifying the status of training data—private leverage will continue to define the terms of compliance, allowing a quasi-legal regime to solidify outside the reach of courts and legislatures.