

**CONTRACTUAL OVERRIDE:  
HOW PRIVATE CONTRACTS UNDERMINE THE GOALS  
OF THE COPYRIGHT ACT FOR LIBRARIES AND RESEARCHERS,  
AND WHAT WE CAN DO ABOUT IT**

by DAVE HANSEN,<sup>1</sup> YUANXIAO XU,<sup>2</sup> and RACHAEL G. SAMBERG<sup>3</sup>

*A wide variety of scholarly and academic uses of copyrighted materials are governed not by copyright law itself but by licenses, terms of service, and other privately crafted contractual terms. In many cases, those terms purport to override exceptions and limitations granted by Congress in the Copyright Act for the benefit of users. As compared to other jurisdictions, the US does not have clear statutory provisions preventing private contracts from overriding certain user rights—rights that are meant to support innovation, teaching, research, and preservation, and designed to strike a careful balance between the interests of the public and copyright owners. Allowing contracts to upset this balance risks granting copyright owners excessive control at the public's expense, ultimately stifling innovation, creativity and the free expression rights of subsequent authors. This paper is about the harm caused by contractual override to two of the most vulnerable and impacted user groups—academic researchers and libraries, and ways to limit that harm.*

INTRODUCTION.....	676
I. LICENSES FOR RESEARCH LIBRARIES.....	687
A. For whom and what do libraries sign agreements?.....	687
B. TDM and AI clauses that override fair use.....	694
C. The knowledge economy.....	703
II. CONSUMER MARKET MATERIALS.....	705
A. The market for consumer works.....	706
B. How consumer contracts limit lawful use for libraries.....	708
1. Redefining the universe of acceptable use: blanket prohibitions on reproduction, adaptation, and distribution.....	709
2. Restricting any use that is not personal and non-commercial.....	712
3. Prohibitions on public performance.....	715
4. Restrictions on circumventing TPMs.....	717
C. Unwillingness to accommodate library and research interests for consumer market materials.....	720
III. PUBLIC WEBSITES AND TERMS OF SERVICE.....	720
A. Examples of uses that benefit the public, and ToS that prohibit these uses.....	721
B. The legality of the prohibitive ToS.....	727
C. Recent cases, failed defenses, and undesirable consequences.....	728
IV. SOLUTIONS.....	729
A. Legislative and Regulatory Approaches.....	730
1. Fulsome Federal Solution: The European Union Framework.....	730

<sup>1</sup> J.D., MSLS, Executive Director, Authors Alliance.

<sup>2</sup> J.D., Staff Attorney, Authors Alliance.

<sup>3</sup> J.D., MLIS. Director, Scholarly Communication & Information Policy, UC Berkeley.

2. Section 108 reform for preservation and interlibrary lending.....	733
3. Federal anti-SLAPP to disincentivize threats against libraries.....	734
4. State legislative solutions.....	735
B. Private Ordering.....	738
1. Expanding library licensing expertise and education.....	738
2. Efficient breach.....	742
C. Judicial approaches.....	744
1. Challenging the validity of the contract.....	745
2. Copyright preemption challenges.....	748
a. Express preemption challenges.....	748
b. Conflict preemption challenges.....	752
3. Other judicial strategies and defenses.....	754
CONCLUSION.....	758

## INTRODUCTION

In February 2025, UC Berkeley Library, Authors Alliance, and the Association of Research Libraries co-hosted a workshop that aimed to tackle the problem of copyright owners and licensing intermediaries using contractual terms to encroach on libraries' and academic researchers' freedom to exercise copyright exceptions and limitations in the United States. This paper accounts for the issues experts discussed at that workshop, specifically: the kinds of agreements that libraries and researchers encounter when using digital content, the particular ways those agreements impinge upon or override underlying copyright exceptions and limitations, and the solutions that may be available based on the type of agreement governing the content in use.<sup>4</sup>

Academic libraries understand the importance of copyright owners' exclusive rights of reproduction, distribution, public display, public performance, and the creation of derivative works.<sup>5</sup> As the U.S. Copyright Office recognized, libraries engage with and uphold the parameters of these exclusive rights to support the research and scholarship that occurs in educational institutions—and university libraries lead the way in educating scholars about compliance with copyright owners' rights under Section 106.<sup>6</sup>

<sup>4</sup> We are thankful to workshop participants Aaron Perzanowski, Danette Pachtner, David Bamman, Emily Sherwood, Eric Harbeson, Giesele Tanasse, Guy Rub, Jack Bernard, Jonathan Band, Katherine Klosek, Kathleen DeLaurenti, Katie Fortney, Katie Zimmerman, Kyle Courtney, Lila Bailey, Maggie Chon, Mark Lemley, Marta Belcher, Mary LaFrance, Miranda Bennett, Mitch Stoltz, Molly Van Houweling, Pamela Samuelson, Rochelle Lundy, and Tim Vollmer for their insights. We are also grateful to Samantha Teremi for her licensing assistance. This paper reflects a good faith interpretation of the discussion at the workshop, supplemented with our additional analysis. Any use of the word “we” with respect to opinions herein refers solely to the views of the authors, and all errors or omissions are our own.

<sup>5</sup> 17 U.S.C. §106. For more on libraries' foundational professional commitment to honoring the rights of copyright owners, *see, e.g.*, ALA Code of Ethics, AM. LIBRARY ASS'N, <https://www.ala.org/tools/ethics> (stating that “We respect intellectual property rights and advocate balance between the interests of information users and rights holders”).

<sup>6</sup> The U.S. Copyright Office had been asked to “study the extent to which copyright owners are experiencing infringements by state entities without adequate remedies under state law” and to

At the same time, Section 106 is by its own terms “subject to Sections 107 through 122,” which incorporate important limitations that balance copyright owners’ rights with those of users.<sup>7</sup> Libraries and scholars depend on these limitations on exclusivity, particularly Section 107 (Fair Use) and Section 108 (Reproduction by Libraries and Archives), to engage in a wide variety of research and educational uses.<sup>8</sup> Other important limitations on copyright owners’ exclusive rights include the freedom to reuse facts and ideas embodied in copyrightable works, the freedom to use public domain works, and the right to teach with copyright-protected materials.<sup>9</sup> For example, when a faculty member at a university wants to download a chapter from a library-licensed e-book to review that chapter for research, the faculty member would need to exercise the right of reproduction. In order for the faculty member to proceed with the reproduction, the act would need to fall within an exception or limitation like Fair Use. When a graduate student wants to examine parole board decision-making across hundreds of board decisions, the student will seek to download those decisions from the parole board’s public website—again implicating the exclusive reproduction right unless a limitation like Fair Use applies. When a library stewards a unique collection item needed by a scholar elsewhere, it depends on Section 108 for interlibrary loan. When a researcher wants to extract and analyze historic housing or agricultural pricing data, she is relying on the uncopyrightable status of such statistics.

Although the limitation placed on copyright terms and copyrightable subject matter, as well as exceptions and limitations (like Section 107 and 108), are statutorily prescribed to support these kinds of research and educational activities, in the modern digital content landscape, they are often being overridden by contracts governing the

---

“consider the extent to which such infringements appear to be based on intentional or reckless conduct.” The Office “heard from representatives of a number of state entities, primarily universities and libraries, who provided information about their efforts to educate their communities about copyright through the adoption of policies and educational programs”—and the Copyright Office concluded that there was not sufficient evidence of unconstitutional conduct or infringement by state universities and libraries; “To the contrary, we received considerable evidence of policies and programs adopted by a number of state entities in order to deter infringement.” U.S. COPYRIGHT OFFICE, COPYRIGHT AND STATE SOVEREIGN IMMUNITY: A REPORT TO THE REGISTER OF COPYRIGHTS, 72 (2021), <https://www.copyright.gov/policy/state-sovereign-immunity/Sovereign%20Immunity%20Report%20final.pdf>. Within the University of California system, for example, and in addition to libraries undertaking copyright education, “there are multiple layers of policies relating to copyright—some are system-wide, some campus-wide, and some operate at a departmental level.” *Id.* at 55-58 (detailing extensive educational outreach by libraries on copyright issues).

<sup>7</sup> 17 U.S.C. § 106.

<sup>8</sup> 17 U.S.C. §§ 107-08. While other exceptions and limitations exist under the Copyright Act (e.g. 17 USC § 110), this paper focuses predominantly on Sections 107 and 108, as they are most impacted by the kinds of contracts discussed herein.

<sup>9</sup> 17 U.S.C. § 102 (facts and ideas not protected by copyright); 17 U.S.C. § 105 (U.S. Government works not protected by copyright); 17 U.S.C. § 110 (limitations on exclusivity for performance and display in a classroom); 17 U.S.C. §§ 302-305 (works enter public domain after duration of copyright expires)..

content that a scholar or student wants to use.<sup>10</sup> Use of that e-book by the above faculty member, or use of that parole board website content by the graduate student, is likely subject (respectively) to a complex license agreement negotiated by the library at that institution or the website's terms of service. And in many cases, as we demonstrate herein, the provisions in these agreements either entirely prohibit or severely limit a user's ability to engage in uses that would otherwise be allowed by limitations such as Section 107 and 108.<sup>11</sup> These contracts may also curb or prevent the lawful use of public domain materials<sup>12</sup> or factual aspects of materials not even protected by copyright.<sup>13</sup>

These access and usage provisions in effect impose a set of contractual restrictions that extend copyright holders' Section 106 monopoly while eliminating the ability for users to rely on the exceptions and limitations that Congress created to balance such exclusivity. Yet, libraries and researchers who violate such terms face a variety of consequences, from threats of litigation and the potential liability for breach of contract, to loss of ongoing access to important research materials for which there is no alternative.<sup>14</sup> Publishers send demand letters threatening damages if a library does not

<sup>10</sup> See, e.g., Stacey L. Bowers, *The Impact of License Agreements on Access to Information: A Challenge for Academic Libraries*, in *COMPUT. NETWORKS, POL'YS & APPLICATIONS* 84 (C.M. Berger & A.D. Streusse eds., 2011) ("[M]any license agreements also restrict or eliminate the library's ability to rely on Sections 107 and 108 of the copyright law. As a result, the library gives up control to provide the services needed and often demanded by its patrons. This loss can deter the library from its stated mission of providing access to information. Once the library has entered into the license agreement, it cannot rely on copyright law because it provides broader rights. The library must abide by the language of the license agreement, no matter how severely such language restricts the library in its ability to engage in fair use and interlibrary loan transactions, and ultimately provide the access it has promised."); Int'l Fed'n of Library Ass'ns & Institutions, *IFLA Statement on Contract Override* (Aug. 16, 2025), <https://repository.ifla.org/items/17cd3348-50c8-4e4f-9af0-8a2bb10321b0> ("[T]he fact that digital materials are typically accessed under licence agreements too often requires libraries to waive the rights they would usually have with physical ownership, both on their own behalf and that of their users. As a result, the space for libraries to work that exists under traditional limitations and exceptions can be lost in favour of minimal access. This includes possibilities to lend, preserve and provide enduring access - activities that are central to their role as long-term stewards of knowledge.").

<sup>11</sup> 17 U.S.C. §§ 107-22. See also Pamela Samuelson, *Justifications for Copyright Limitations & Exceptions*, in *COPYRIGHT LAW IN AN AGE OF LIMITATIONS AND EXCEPTIONS* (Ruth Okediji ed., 2017), <https://ssrn.com/abstract=2476669>.

<sup>12</sup> U.S. CONST. U.S. Const. art. I, § 8, cl. 8. See also 17 U.S.C. §§ 302-305.

<sup>13</sup> 17 U.S.C. § 102(b) ("In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work."); see also *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003) (Copyright Act's denial of protection to "ideas" is a "First Amendment accommodation" "built-in" to the Act).

<sup>14</sup> COMMUNIA, *Unfair Licensing Practices: The Library Experience* (May 2025), [https://communia-association.org/wp-content/uploads/2025/05/Unfair-licensing-practices\\_the-library-experience.pdf](https://communia-association.org/wp-content/uploads/2025/05/Unfair-licensing-practices_the-library-experience.pdf) ("When such contractual prohibitions are in place, publishers can simply remove the library's access to the licensed content... [P]ublishers can, and have in the past, cut the entire campus' access to the publisher's site as a consequence of a single user's alleged mis-use of the material."). Some publishers have even demanded retraction of published scholarship for

“cure” alleged contractual breaches by scholars whom the publishers claim have not conformed with the library’s license agreement in some way.<sup>15</sup> At a minimum, publishers’ complex restrictions in agreements create confusion for researchers about what is or is not permitted under an agreement, which has a demonstrable chilling effect on research; scholars are unclear about what rights remain, and what types of research activities are permitted.<sup>16</sup>

When these usage agreements contractually bar and impose liability for activities that Sections 107 and 108 or the public domain allow for, then none of the following essential research and educational undertakings with licensed materials is possible:

- Viewing (“public display” right);
- Downloading, printing, and copying reasonable portions (“reproduction” right);
- Preserving, caching, and creating archival/back-up copies (“reproduction” right);
- Dissemination for classroom use and course packs or electronic course reserves (“distribution” right);
- Scholarly sharing of insubstantial amounts to support research or education (“reproduction” and “distribution”);
- Interlibrary loan (“reproduction” and “distribution”); and more.

---

alleged violations of underlying license agreements on which the researchers relied for access.

*See, e.g.,* Adam Marcus, A Very Unfortunate Event: Paper on COVID-19 Vaccine Hesitancy Retracted, RETRACTION WATCH (July 30, 2021),

<https://retractionwatch.com/2021/07/30/a-very-unfortunate-event-paper-on-covid-19-vaccine-hesitancy-retracted/>.

<sup>15</sup> Publishers regularly seek to impose liability on libraries for the acts of their users. *See, e.g.* COMMUNIA, *supra* note 14 (“In recent years, academic libraries have seen the introduction of clauses on liability that expose libraries to unprecedented levels of responsibility for the actions of library users, together with strong penalties for non-compliance. These hostile provisions greatly increase the risk of an institution being sued by a rightholder — as a result, negotiations in some cases have resulted in deadlock because the library (and its parent institution) simply cannot undertake the legal risk, particularly for publicly funded institutions.”); Int’l Fed’n of Library Ass’ns & Institutions, *supra* note 10 (recognizing “[t]he imposition of unlimited liability on libraries for the actions of their users (even when having provided clear and adequate information), as well as for incidents of exceeding licence terms while acting reasonably and in good faith”); FinELib’s Collaboration With ACS Ends, FINELIB (Aug. 20, 2024), <https://finelib.fi/finelibs-collaboration-with-acs-ends/> (FinELib, a consortium of Finnish universities, research institutions, and public libraries, recently failed to reach agreement with the American Chemical Society due to “fundamental differences of views on the liability issues between the subscribing organisations and ACS.”) In one recent draft agreement, satellite data and earth imagery company Planet Labs proposed that, “Licensee is responsible for any and all acts or omissions of its Authorized Users as if it had undertaken such acts or omissions itself,” which UC Berkeley successfully negotiated against. Planet Labs, Master Content License Agreement, [https://drive.google.com/file/d/19QaSCzKtw\\_v8nXY\\_HxJJ4uSx9uQyUEun/view?usp=sharing](https://drive.google.com/file/d/19QaSCzKtw_v8nXY_HxJJ4uSx9uQyUEun/view?usp=sharing).

<sup>16</sup> *Id.*; *see also* Authors Alliance, Text and Data Mining under U.S. Copyright Law: Landscape, Flaws & Recommendations (Oct. 2024), <https://www.authorsalliance.org/wp-content/uploads/2024/11/Text-and-Data-Mining-Report-102024.pdf> (low risk tolerance of researchers and universities in the face of legal uncertainty results in research projects being abandoned).

This hardly seems like it should be possible or lawful, and yet users are regularly faced with contracts that purport to prohibit scholarly downloading, library lending, text data mining (“TDM”), and more.

What is being sacrificed here is the very equilibrium that copyright law aims to strike: the balance between the rights of copyright owners who are granted limited exclusivity over creative works, and the rights of the public to gain access to those works, learn from them, and build upon them.<sup>17</sup> Because rigid exclusivity can “stifle the very creativity which that law is designed to foster,”<sup>18</sup> the law has “limited the scope of copyright protection to ensure that a copyright owner’s monopoly does not harm the

---

<sup>17</sup> L. RAY PATTERSON & STANLEY W. LINDBERG, *THE NATURE OF COPYRIGHT: A LAW OF USERS’ RIGHTS* (1991); *Twentieth Century Music Corp. v. Aiken*, 422 U.S. 151, 156 (1975) (“The limited scope of the copyright holder’s statutory monopoly, like the limited copyright duration required by the Constitution, reflects a balance of competing claims upon the public interest: Creative work is to be encouraged and rewarded, but private motivation must ultimately serve the cause of promoting broad public availability of literature, music, and the other arts. The immediate effect of our copyright law is to secure a fair return for an ‘author’s’ creative labor. But the ultimate aim is, by this incentive, to stimulate artistic creativity for the general public good.”); *United States v. Paramount Pictures*, 334 U.S. 131, 158 (1948) (“‘The sole interest of the United States and the primary object in conferring the monopoly lie in the general benefits derived by the public from the labors of authors.’ It is said that reward to the author or artist serves to induce release to the public of the products of his creative genius”) (*quoting* *Fox Film Corp. v. Doyal*, 286 U.S. 123 (1932)). *See also* REGISTER OF COPYRIGHTS, REPORT ON THE GENERAL REVISION OF THE U.S. COPYRIGHT LAW, 87TH CONG. (COMM. PRINT 1961) (“As reflected in the Constitution, the ultimate purpose of copyright legislation is to foster the growth of learning and culture for the public welfare, and the grant of exclusive rights to authors for a limited time is a means to that end. A fuller statement of these principles was contained in the legislative report [H.R. REP. NO. 60-2222 (1908)] on the Copyright Act of 1909: ‘The enactment of copyright legislation by Congress under the terms of the Constitution is not based upon any natural right that the author has in his writings, for the Supreme Court has held that such rights as he has are purely statutory rights, but upon the ground that the welfare of the public will be served and progress of science and useful arts will be promoted by securing to authors for limited periods the exclusive rights to their writings. The Constitution does not establish copyrights, but provides that Congress shall have the power to grant such rights if it thinks best. Not primarily for the benefit of the author, but primarily for the benefit of the public, such rights are given. Not that any particular class of citizens, however worthy, may benefit, but because the policy is believed to be for the benefit of the great body of people, in that it will stimulate writing and invention to give some bonus to authors and inventors. In enacting a copyright law Congress must consider two questions: First, how much will the legislation stimulate the producer and so benefit the public, and, second, how much will the monopoly granted be detrimental to the public? The granting of such exclusive rights, under the proper terms and conditions, confers a benefit upon the public that outweighs the evils of the temporary monopoly.’”). This balance has existed since the first copyright law, Statute of Anne. *See* Jessica Litman, *Readers’ Copyright*, 8 J. COPYRIGHT SOC’Y U.S. 325, 334 (2011) (“The first copyright statute gave copyright owners narrowly bounded rights. It allowed readers the freedom to engage in a host of potentially valuable uses outside of the boundaries.”).

<sup>18</sup> *Stewart v. Abend*, 495 U.S. 207, 236 (1990).

public interest.”<sup>19</sup> Despite these important built-in limits on copyright owners’ reach—and the fact that these limitations reflect the bargain between the copyright owners and the public—private contracts are now reaching for excessive control benefiting copyright owners at the expense of the public.<sup>20</sup> As Professor Jessica Litman warns, “Copyright owners may not always be wise in the uses they choose to permit or prohibit; their immediate goals may or may not align with society’s overarching goals for having a copyright system in the first place. Readers deserve guarantees that their liberties will not be subject to copyright owner whims.”<sup>21</sup>

The contracts discussed in this paper are being deployed to prevent researchers from making lawful adaptations or transformative uses under fair use;<sup>22</sup> and to prohibit libraries from creating preservation copies of e-books and other digital media that are at risk of disappearing<sup>23</sup>—even though such efforts are crucial for advancing science and

<sup>19</sup> Google LLC v. Oracle Am., Inc., 593 U.S. 1, 2 (2021). These limitations are also what makes copyright law compliant with the First Amendment. See Eldred v. Ashcroft, 537 U.S. 186, 219-20 (2003) (describing Fair Use and other limitations on copyright as built in First Amendment safeguards).

<sup>20</sup> JESSICA LITMAN, DIGITAL COPYRIGHT 15 (rev. ed. 2017) (“United States copyright law has until now divided up the value that inheres in works of authorship to permit authors (and their employers and publishers) to control and therefore profit from some uses of their works, while forbidding them from controlling others. Authors are given enough control to enable them to exploit their creations, while not so much that consumers and later authors are unable to benefit from the protected works.”).

<sup>21</sup> Jessica Litman, *Readers’ Copyright*, 58 J. COPYRIGHT SOC’Y. U.S.A 325 (2011).

<sup>22</sup> For illustration purposes in this paper, we have compiled several agreements presented by publishers or their licensing intermediaries to UC Berkeley Library. We attach relevant parts of these agreements either in their pre-negotiated form (i.e. as received from the publisher or vendor) or as executed. Within the agreements we have compiled, to view sample contractual terms prohibiting adaptation, modification, or the creation of derivative works, see, e.g., Moody’s Analytics, Terms of Agreement, [https://drive.google.com/file/d/1Gy3WiL8r8BBxul5hNVOjd29VI\\_dSrYb5/view?usp=sharing](https://drive.google.com/file/d/1Gy3WiL8r8BBxul5hNVOjd29VI_dSrYb5/view?usp=sharing) (Preamble); Planet Labs, Master Content License Agreement, [https://drive.google.com/file/d/19QaSCzKtw\\_v8nXY\\_HxJJ4uSx9uQyUEun/view?usp=sharing](https://drive.google.com/file/d/19QaSCzKtw_v8nXY_HxJJ4uSx9uQyUEun/view?usp=sharing) (Par. 3.2(c)); Clarivate, Terms Master Client Agreement & Product Service Terms Addendum, [https://drive.google.com/file/d/152pS-H9NHsfPw9xtdMkIbF947lGUvXLT/view?usp=drive\\_link](https://drive.google.com/file/d/152pS-H9NHsfPw9xtdMkIbF947lGUvXLT/view?usp=drive_link) (Par. 2(g)).

<sup>23</sup> Glyn Moody, *We Risk Losing Access to the World’s Academic Knowledge, and Copyright Makes Things Worse*, WALLED CULTURE (Mar. 27, 2024), <https://walledculture.org/we-risk-losing-access-to-the-worlds-academic-knowledge-and-copyright-makes-things-worse/> (last visited June 6, 2025); Michelle M. Wu, *Restoring the Balance of Copyright: Antitrust, Misuse, and Other Possible Paths to Challenge Inequitable Licensing Practices* (2021), at 13 (“By limiting what can be loaned through interlibrary loan, all communities face an unnecessary loss of access. By preventing preservation, there is no reliable way to ensure that the knowledge of this generation will be preserved for future users...Across the board, every limitation will harm poorer communities more than they will wealthy ones and future generations more than current ones, increasing the nation’s existing inequality divide...[S]ociety will only be able to retain access to works as long as they can afford to pay repeatedly for the same content; and there is no guarantee that any information available through licensing will exist beyond today.”)



maintaining cultural and historical records, and are expressly authorized under the Copyright Act. Researchers engaging in text and data mining (“TDM”)—a critical method for advancing knowledge in fields ranging from the humanities to medicine, and one courts have long deemed to be Fair Use—may find themselves unable to analyze digital works due to restrictive licensing terms.<sup>24</sup> The license agreements that libraries sign to secure access and usage rights for campus, the online terms of service governing the content on a website a scholar wants to use, and the end-user license agreement that a scholar signs individually for use of an e-book all regularly prohibit each of these acts.

These contractual restrictions can also extend beyond copyright to exert control over data or ideas within or about the works, enabling copyright owners to extend limitations on access and use to unprotected elements that were traditionally free for the public to build upon.<sup>25</sup> Likewise, contracts are used to impose copyright-like restrictions on use of materials that have long entered the public domain, recapturing and extending an effective monopoly over use of works that, constitutionally, are only supposed to be protected for limited times.<sup>26</sup> In each of these cases, contractual override undermines the fundamental balance of copyright law, shifting away from support of uses that benefit the public and aid new creation and toward private copyright owners in ways that Framers of the Constitution never intended.<sup>27</sup>

How did we get here? From the earliest days of the web, scholars identified the challenges that could arise if contracts replaced the default rules of copyright at a massive scale, through one-to-one relationships between online distributors and consumers.<sup>28</sup> Even those who have supported strong copyright control have recognized the power that contracts might hold for copyright owners. For example, in 1993, law professor Jane Ginsburg said this:

<sup>24</sup> Rachael Samberg & Dave Hansen, *Restricting Innovation: How Publisher Contracts Undermine Scholarly AI Research*, AUTHORS ALL. (Dec. 6, 2024), <https://www.authorsalliance.org/2024/12/06/restricting-innovation-how-publisher-contracts-undermine-scholarly-ai-research/>.

<sup>25</sup> Niva Elkin-Koren, *Navigating the Copyright/Contract Interface in the Generative AI Era*, 39 BERKELEY TECH. L. J. 1137, 1142, n.23 (2024).

<sup>26</sup> Derek Fincham, *Bridging Copyright and Cultural Heritage with the Public Domain*, 92 UMKC L. REV. 263, 274–75 (2023) (discussing how libraries and museums impose usage restrictions on exact scans of public domain works despite the *Bridgeman* [Bridgeman Art Library, Ltd. v. Corel Corp., 25 F. Supp. 2d 421 (S.D.N.Y. 1998), *reconsidered* in *Bridgeman Art Library, Ltd. v. Corel Corp.* 36 F. Supp. 2d 191 (S.D.N.Y. 1999)] holding that such reproductions lack sufficient originality for copyright protection); *see also* Andrea Wallace & Ellen Euler, *Revisiting Access to Cultural Heritage in the Public Domain: EU and International Developments*, 51 IIC 823 (2020).

<sup>27</sup> *See* Pamela Samuelson, *Copyright and Freedom of Expression in Historical Perspective*, 10 J. INTELL. PROP. L. 319, 334–35 (2003) (describing the shift from private ordering and enforcement through stationers in the “pre-modern” era, to positive law enacted by Congress and enforced in the courts, and now the shift back toward “post-modern” copyright dominated by private agreements and self-policed enforcement).

<sup>28</sup> *See, e.g.*, Julie Cohen, *Copyright and the Jurisprudence of Self Help*, 13 BERKELEY TECH. L. J. 1089 (1998) (“Courts may, and should, apply principles of preemption and freedom of speech to invalidate license terms authorized by Article 2B that are inconsistent with copyright limitations.”); Niva Elkin-Koren, *Copyright and the Limits of Freedom of Contract*, 12 BERKELEY TECH. L. J. 93 (1997).



In the digital environment posited here, contract protection may not be the fragile creature presumed in prior intellectual property preemption decisions. If access to works could be obtained only through the information provider (directly or through an authorized online distributor), and if copying could be electronically tracked or prevented, no ‘third parties’ to the contract would exist. When ‘we’re all connected,’ no functional difference may exist between a contract and a property right. At that point, it becomes necessary to consider whether limitations incorporated in the copyright law should be imported to its contractual substitute.<sup>29</sup>

Ginsburg’s description became reality through a few connected legal developments over the next two decades. First, some courts accept attempts by distributors to define how users can interact with copyrighted works by licenses rather than by a default set of rules provided by the Copyright Act. This included a rejection of a “digital first sale” right—which could have granted purchasers of digital copies a set of rights analogous to those who buy physical goods, such as the right to sell, lend, or otherwise dispose of that copy.<sup>30</sup> In the absence of such a right, the rights of an owner are typically delineated by the licenses from the copyright owner.<sup>31</sup> For example, a library that seeks to lend e-books it purchases to its users must follow the licensing terms of the e-books. Similarly, licensing terms can prevent users from taking what they have purchased from one proprietary platform to another, or sharing their copies with another account—even though traditionally first sale doctrine has allowed for lending to friends as well as reselling one’s purchased copies. These contractual restrictions are further reinforced by

<sup>29</sup> Jane C. Ginsburg, *Copyright Without Walls?: Speculations on Literary Property in the Library of the Future*, 42 REPRESENTATIONS 53 (1993).

<sup>30</sup> U.S. DEP’T OF COMMERCE, INFORMATION INFRASTRUCTURE TASK FORCE, INTELLECTUAL PROPERTY AND NATIONAL INFRASTRUCTURE: THE REPORT OF THE WORKING GROUP ON INTELLECTUAL PROPERTY RIGHTS, 10-12, 177-78, 230 (1995); U.S. COPYRIGHT OFFICE, DMCA SECTION 104 REPORT: A REPORT OF THE REGISTER OF COPYRIGHTS PURSUANT TO § 104 OF THE DIGITAL MILLENNIUM COPYRIGHT ACT (2001); U.S. PATENT & TRADEMARK OFFICE, WHITE PAPER ON REMIXES, FIRST SALE, AND STATUTORY DAMAGES: COPYRIGHT POLICY, CREATIVITY, AND INNOVATION IN THE DIGITAL ECONOMY (2016) (all rejecting digital first sale). See also Matthew Chiarizio, *An American Tragedy: E-Books, Licenses, and the End of Public Lending Libraries?*, 66 VANDERBILT L. REV. 615 (2013) (detailing history of rejection of first sale for licenses); Aaron Perzanowski & Jason Schultz, *Digital Exhaustion*, 58 UCLA L. REV. 889 (2011); Kanchana Kariyawasam & Royal Raj Subburaj, *Importance of the doctrine of digital exhaustion in copyright law*, 33 INT’L J. L. INFO. TECH. 1 (2025); Michelle M. Wu, *Restoring the Balance of Copyright: Antitrust, Misuse, and Other Possible Paths to Challenge Inequitable Licensing Practices* (2021),

<https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=3430&context=facpub>, at 11

(“One relatively recent development has upset this balance, aiming to remove traditionally copyright-based transactions from the scrutiny of the courts. That development is the use of contract to abrogate the rights that normally attach with acquiring a copy of a copyrighted work. By licensing a work instead of selling it, no transfer of a copy is ever made, and therefore, the copyright interests formerly protected by ownership and the right of alienation are replaced by the rights of a renter against a landlord. The exceptions built into the copyright code ... are therefore avoided through contract.”)

<sup>31</sup> See Section I.A., *infra*.

technological protection measures (TPMs), which can prevent even lawful downstream uses (such as preservation or format-shifting) at the technical level.<sup>32</sup>

Second, the rise of mass-market end-user license agreements (EULAs) and online terms of service (“ToS”, alternatively called “terms of use”) further entrenched contractual control over digital works.<sup>33</sup> These agreements, often presented to academic researchers as non-negotiable “click-through” contracts (and, in the case of many websites, “browse-wraps” mandating consent merely through use of the site), have become ubiquitous.<sup>34</sup> Unlike the statutory framework which grants users certain rights—such as fair use and first sale, EULAs and ToS often explicitly restrict these rights, dictating how, when, and if a given user can engage with a digital work, as we discuss in Section II and III below.

*ProCD v. Zeidenberg*<sup>35</sup> established a rule that influenced many subsequent court decisions that a user can validly accept the terms of a contract by their conduct, even in cases where access or ability to understand the terms is limited.<sup>36</sup> Moreover, the *ProCD* court held, as a matter of principle, contract terms will not be held preempted by federal copyright law, that “a simple two-party contract is not ‘equivalent to any of the exclusive rights within the general scope of copyright’ and therefore may be enforced.” Several circuits still doggedly follow the ruling in *ProCD* with regard to copyright preemption, allowing plaintiffs to state a copyright-adjacent contract claim,<sup>37</sup> though these contracts may limit users’ access to rights such as fair use. While the prevailing view is that these contracts are enforceable, there is now some hope to challenge at least some of them in court, as we discuss in Part IV. C. of this paper below.<sup>38</sup>

Finally, a third development in the marketplace has made contractual override far more consequential than it was in the past. Until more recently, the real-world impact of these legal developments has been tempered by the ability of users to obtain copyrighted

<sup>32</sup> DMCA 1201.

<sup>33</sup> Pamela Samuelson, *Possible Futures of Fair Use*, 90 WASH. L. REV. 815 (“Two techniques commonly used to impede Fair Uses of copyrighted works are the adoption of mass-market license agreements containing restrictive terms and the implementation of technical protection measures (TPMs) that prevent access to or copying of digital works.”)

<sup>34</sup> Amit Elazari Bar On, *Unconscionability 2.0 and the IP Boilerplate: A Revised Doctrine of Unconscionability for the Information Age*, 34 BERKELEY TECH. L.J. 567, 630–36 (2019).

<sup>35</sup> 86 F.3d 1447, 1453 (7th Cir. 1996).

<sup>36</sup> See, e.g., *Register.com v. Verio*, 356 F.3d 393 (2d Cir. 2004) (contract terms were held to be binding despite not appearing until the transaction was concluded); *Derriman v. Mizzen and Main LLC*, 2023 WL 9022723 (M.D. Fla. Dec. 29, 2023) (finding the terms binding when hyperlinked in small underlined fonts next to a clickable button saying “GET 15% OFF”).

<sup>37</sup> The 5th, 11th and Fed Cir are strictly following *ProCD*, categorically holding that contract claims are never preempted by copyright law. See *Lipscher v. LRP Publ’ns, Inc.*, 266 F.3d 1305, 1318 (11th Cir. 2001) (holding that “claims involving two-party contracts are not preempted because contracts do not create exclusive rights,” even where the relevant agreement “deal[s] only with copying and redistribution of the materials.”); *Bowers v. Baystate Techs., Inc.*, 320 F.3d 1317, 1325 (Fed. Cir. 2003); *Taquino v. Teledyne Monarch Rubber*, 893 F.2d 1488 (5th Cir. 1990).

<sup>38</sup> *In re Jackson*, 972 F. 3d 25 (2020); *X Corp. v. Bright Data Ltd.*, 733 F. Supp. 3d 832 (N.D. Cal. 2024). And with regard to copyright preemption, there are several circuits criticizing the rigid approach taken by *ProCD*, a trend first set by *Wrench LLC v. Taco Bell Corp.*, 256 F.3d 446, 457 (6th Cir. 2001). See also *ML Genius v. Google*, 2022 WL 710744 (2d Cir. Mar. 10, 2022).

works in other formats: most streaming films had DVD alternatives; most e-books had physical versions available, too, where no enforceable restrictive terms are attached to a purchase. In recent years, however, copies not subject to licenses have increasingly become unavailable.<sup>39</sup> Libraries and researchers, for example, no longer have the ability to walk into a bookstore to buy the corpus of literature they need for their work; their access to the works is now entirely subject to the whims of licensors in the digital world.<sup>40</sup>

Some of this “born digital” content (as we discuss in Part II. B. *infra*) is available not only with restrictive licenses, but also bundled behind TPMs. Proprietary TPMs may tie an e-book’s usability to a specific platform or software, meaning the content could vanish if the platform becomes obsolete or the provider goes out of business, introducing external dependencies that jeopardize long-term access. Libraries, left with no way to make archival or accessible copies, are effectively shut out of their preservation mission. This threatens not only future access to our cultural and scholarly record but also the ability of researchers—especially those who rely on assistive technologies—to engage with digital collections.<sup>41</sup>

While contractual override affects libraries and researchers through a wide range of contract types and use cases, the particular rights and limitations problems it causes—and the solutions these problems call for—differ across at least three areas.

In *Part I*, we examine the licenses that research libraries negotiate and sign for content like books, journals, data, and films that scholars seek to use. We explore the pitfalls that libraries encounter in trying to preserve fair uses and library exceptions in the wake of libraries’ diminished bargaining power relative to publishers and vendors. In some cases, the override may derive from copyright owners’ unfamiliarity with how scholars need to engage with resources, but predominantly the override clauses curb lawful uses either based on publishers’ fears or as their means of occupying the market for a type of use.

*Part II* explores scholars’ need for “mass market” materials in their research and education. “Mass market” materials include books created for a general audience, films and television, along with specific works now created only in a digital form (e.g., Amazon e-books, Netflix originals). We address how non-negotiable agreements or

<sup>39</sup> See, e.g., Brad Adgate, *Why Disney Is Cutting Back On DVDs After 'Guardians Of The Galaxy'*, FORBES (Aug. 2, 2023), <https://www.forbes.com/sites/bradadgate/2023/08/02/with-sales-dropping-disney-will-no-longer-sell-dvds-in-australia/>; Laura Carollo, *Physical Movies - Statistics & Facts*, STATISTA (Sept. 10, 2024), <https://www.statista.com/topics/12707/physical-movies/#topicOverview> (market data on the collapse of physical video market); DIGITAL ENTERTAINMENT GROUP, YEAR-END 2024 DIGITAL MEDIA ENTERTAINMENT REPORT (Feb. 14, 2025), [https://www.degonline.org/portfolio\\_page/deg-q4-2024-digital-media-entertainment-report/](https://www.degonline.org/portfolio_page/deg-q4-2024-digital-media-entertainment-report/) (“Spending on physical formats, including DVD, Blu-ray and 4K UHD continued to decline, dropping 23.4 percent for full year 2024 and the fourth quarter.”)

<sup>40</sup> COMMUNIA, *supra* note 14.

<sup>41</sup> University of Michigan Library’s Comment to Mandatory Deposit of Electronic-Only Books Proposed Rulemaking [Docket No. 2016–03] <https://www.regulations.gov/comment/COLC-2018-0002-0015>

EULAs for these materials often ban uses deemed fair or permissible under exceptions and limitations like Sections 107 and 108.

*Part III* turns to content that scholars use for research that is available on publicly accessible websites (e.g. Twitter, Facebook, LinkedIn, Reddit). In this section, we address the Terms of Service of publicly accessible websites that similarly attempt to override scholars' rights, and what potential role the library plays in this ecosystem.

After describing these three key types of agreements affecting library missions and services, and addressing their impact on researchers, *Part IV* analyzes legislative, private ordering, and judicial solutions to resolve challenges caused by contractual override in these contexts. We begin by accounting for our gold standard result: Federal legislation, akin to that in the European Union, that protects scholarly research and library exceptions from contractual override. While Congress briefly considered one version of this (more than twenty years ago), foreign jurisdictions have taken the lead and offer apt guides for future efforts protecting scholars and libraries.<sup>42</sup> Notably, most countries in Europe have laws safeguarding key exceptions and limitations related to research, so that private companies cannot derail the progress of science and culture by asserting excessive control over their copyrighted materials.<sup>43</sup> Developments, such as in the UK in 2014,<sup>44</sup> where licensing terms are considered void if they are contrary to copyright exceptions and limitations,<sup>45</sup> give us hope that, even in a jurisdiction that highly values the freedom of contract, we can still achieve meaningful changes to protect copyright exceptions and limitations. Yet the pathway to such legislation is challenging. State legislation holds some promise with respect to certain aspects of library-negotiated agreements, but must tread carefully to avoid federal encroachment.

In the absence of comprehensive federal or state legislation, we find private ordering—in the form of extensive education—to be the most promising for addressing

---

<sup>42</sup> See Jonathan Band, *Protecting User Rights Against Contract Override* (PIJIP/TLS Rsch. Paper Series No. 97, 2023), <https://digitalcommons.wcl.american.edu/research/97/> (quoting provisions in copyright law around the world that protect limitations and exceptions from override by contract).

<sup>43</sup> *Id.* For theory underpinning these approaches, see, e.g.: WIPO STANDING COMMITTEE ON COPYRIGHT AND RELATED RIGHTS, SCCR/34/CHART ON LIBRARIES AND ARCHIVES ("Limitations and exceptions granted by law should not be emptied of their effect through the application of technological protection measures. Libraries should be allowed to acquire and apply tools to remove such measures, in order to fulfil [*sic*] their public interest missions."); *Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired or Otherwise Print Disabled* art. 7, June 27, 2013, 52 I.L.M. 1312 ("Contracting Parties shall take appropriate measures, as necessary, to ensure that when they provide adequate legal protection and effective legal remedies against the circumvention of effective technological measures, this legal protection does not prevent beneficiary persons from enjoying the limitations and exceptions provided for in this Treaty").

<sup>44</sup> *Copyright, Designs and Patents Act 1988*, c. 48 (U.K.).

<sup>45</sup> *Id.* See in particular: Section 29 (Research, private study and text and data analysis for non-commercial research) "To the extent that a term of a contract purports to prevent or restrict the doing of any act which, by virtue of this section, would not infringe copyright, that term is unenforceable"; Section 29A (Copies for text and data analysis for non-commercial research): "To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable."

override in library license agreements; and, we find judicial solutions—in the form of conflict preemption—to hold merit with respect to some public website ToS and certain EULAs. Given the trajectory of research override provisions in the digital era, if Congress, state legislatures, libraries, and academic institutions do not engage critically with these solutions, then U.S. education, research, and innovation may be irreparably constrained, and the public benefit of the copyright bargain lost.<sup>46</sup>

## I. LICENSES FOR RESEARCH LIBRARIES

### A. For whom and what do libraries sign agreements?

Academic library users include the students, researchers, and faculty for whom libraries license digital content like e-books, journals, databases, data, and more. These academic users read, study, analyze, and incorporate understandings from these materials in their work in the classroom, their research, and their scholarship. To understand how some publishers and vendors are using these library license agreements to override copyright exceptions, it is helpful first to establish what these library license agreements cover, and the impact they have on research, innovation, and education.

The University of California may be illustrative for context. Nearly five billion dollars each year is invested to support the research conducted at the University of California.<sup>47</sup> The outputs of this research—in the form of inventions, scholarship, policy-setting, and more—better the world, and advance knowledge in ways that directly improve health, education, technology, literature, art, and quality of life.<sup>48</sup> With respect to scholarly outputs, University of California research results in approximately 10% of all scholarly publishing in the U.S.<sup>49</sup> This research is performed by faculty, graduate and undergraduate students, postdoctoral researchers, and professional staff—all of whom rely on access to a broad array of academic, cultural, and creative

<sup>46</sup> U.S. CONST. art. I, § 8, cl. 8.; *Google LLC v. Oracle America, Inc.*, 141 S.Ct. 1183, 1186, 593 U.S. 1, 2 (U.S., 2021) (“Because such exclusivity may trigger negative consequences, Congress and the courts have limited the scope of copyright protection to ensure that a copyright holder’s monopoly does not harm the public interest”); *Authors Guild v. Google*, 804 F.3d 202, 214 (2d Cir. 2015) (“copyright’s overall objective [is] contributing to public knowledge”).

<sup>47</sup> UNIVERSITY OF CALIFORNIA, ACCOUNTABILITY REPORT: RESEARCH (2023), <https://accountability.universityofcalifornia.edu/2023/chapters/chapter-9.html> (last visited June 5, 2025).

<sup>48</sup> *Impacts of Research*, UNIV. OF CAL., BERKELEY Univ. of Cal., Berkeley, <https://www.berkeley.edu/research-impact/> (last visited June 22, 2025).

<sup>49</sup> UNIVERSITY OF CALIFORNIA, ACCOUNTABILITY REPORT: RESEARCH (2020), <https://accountability.universityofcalifornia.edu/2020/chapters/chapter-9.html#9.2.2> (last visited June 5, 2025); *12 Amazing UC Discoveries of 2024*, UNIV. OF CAL., <https://www.universityofcalifornia.edu/news/12-amazing-uc-discoveries-2024> (last visited June 22, 2025); *Ten Ways UC Has Made California Better*, UNIV. OF CAL., <https://www.universityofcalifornia.edu/news/ten-ways-uc-has-made-california-better> (last visited June 22, 2025).

content to power their work.<sup>50</sup> The electronic resources they utilize include scholarly publications, databases, data, literary works, social media, and more.<sup>51</sup>

While some research content types (like social media posts) are available on public websites, electronic access to the bulk of what scholars need to study is typically secured by campus libraries through institutional license agreements, sometimes referred to as electronic resource agreements or “e-resource” agreements.<sup>52</sup> Libraries enter into e-resource agreements either directly with publishers, or with content aggregators (like Proquest, now a Clarivate company, or EBSCO) who themselves compile and re-license scholarly journals, data, or e-books from multiple publishers.<sup>53</sup> Large e-resource agreements might cover thousands of journal titles by a single publisher, and represent

<sup>50</sup> UNIV. OF CAL. LIBRARY., UNIV. OF CAL. ACCOUNTABILITY REPORT: CHAPTER 9, RESEARCH (2019), <https://accountability.universityofcalifornia.edu/2019/chapters/chapter-9.html> (last visited June 22, 2025); UNIV. OF CAL. LIBR., UNIV. OF CAL. LIBRARY: VISION AND PRIORITIES, <https://libraries.universityofcalifornia.edu/about/vision-and-priorities/> (last visited June 22, 2025);

<sup>51</sup> Regarding the scope and array of content needed, see UNIV. OF CALIFORNIA LIBRARY., UNIV. OF CAL. LIBRARY. / UNIV OF CALIFORNIA OFF. OF THE PRES. (UCOP) STAT.: FISCAL YR. 2023-2025 (2025), [https://libraries.universityofcalifornia.edu/wp-content/uploads/2025/04/UCL--UCOP-Statistics-Fiscal-Year-2023-2024-Public-Summary\\_v2.pdf](https://libraries.universityofcalifornia.edu/wp-content/uploads/2025/04/UCL--UCOP-Statistics-Fiscal-Year-2023-2024-Public-Summary_v2.pdf).

Regarding the types of materials scholars rely on, see, e.g., David Bamman et al., Measuring diversity in Hollywood through the large-scale computational analysis of film, *PROC. NATL. ACAD. SCI. U.S.A.* 121 (46) e2409770121, <https://doi.org/10.1073/pnas.2409770121> (2024) (analyzing films); Stuart Wolpert, *The Teenage Brain on Social Media*, UCLA NEWSROOM (June 1, 2016), <https://newsroom.ucla.edu/releases/the-teenage-brain-on-social-media> (analyzing social media); Kara Mankem, *Study Finds Persistent Spike in Hate Speech on X*, U.C. BERKELEY RSCH. (Feb. 13, 20254), <https://vcresearch.berkeley.edu/news/study-finds-persistent-spike-hate-speech-x> (analyzing social media); Big Data Meets Literary Analysis: Digital Humanities Research at the I School, UC BERKELEY SCH. OF INFO (2018), <https://www.ischool.berkeley.edu/news/2018/big-data-meets-literary-analysis-digital-humanities-research-i-school> (analyzing literary works); Ian Dalke, *I Come Before You a Changed Man: “Insight,” Compliance, and Refurbishing Penal Practice in California*, 49 L. & SOC. INQUIRY (2024) (analyzing data of parole decisions).

<sup>52</sup> See, e.g., Bowers, *supra* note 10 at 79; BECKY ALBITZ, LICENSING & MANAGING ELEC. RESOURCES 51 (2008); LESLEY ELLEN HARRIS

, LICENSING DIGITAL CONTENT: A PRACTICAL GUIDE FOR LIBRARIANS (3d ed. 2018).

<sup>53</sup> See Sue Mattingly, *What Are Journal Content Aggregators?*, WILEY (March 9, 2022), <https://www.wiley.com/en-us/network/publishing/societies/publishing-strategy/what-are-journal-content-aggregators> (“Content aggregators are companies that host content from many different publishers on the aggregator’s platform. [EBSCO] aggregates content from across many scholarly research publishers and collects them into subscription products, which libraries purchase in order to give their users broad-based content access across many publishers in a single interface. These aggregated products are especially valuable to publishers because they help content reach subscribers that otherwise would haven’t [*sic*] the budget or administrative resources to work with each publisher individually – this can be especially true for organizations like community colleges, public libraries, schools, and other secondary markets for scholarly research.”). See also Sharon E. Farb, *Negotiating Use, Persistence, and Archiving: A Study of Academic Library and Publisher Perspectives on Licensing Digital Resources* (2006) (Ph.D. dissertation, University of California, Los Angeles).

millions of individual journal articles being licensed by an institution. For example, each year, the licensing librarian at UC Berkeley reviews and negotiates approximately 125 new e-resource agreements as well as hundreds more contract renewals, which (in combination with additional University of California systemwide agreements) yield the licensing of nearly 4.2 million electronic titles for UC Berkeley scholars to use—a figure that, in 2024, reflected a 23% growth in digital and electronic titles over the previous year.<sup>54</sup>

All of these e-resource agreements represent huge business for publishers, so it is perhaps unsurprising that publishers ultimately wish to highly regulate (and in some cases charge extra) for research rights like Fair Use already afforded by copyright exceptions.<sup>55</sup> To understand the enormous scale of the e-resource licensing industry, consider that the 119 U.S.-based academic research libraries who are currently members of the Association of Research Libraries *alone* spend approximately \$2.4 billion annually on ongoing subscriptions.<sup>56</sup> For their part, across the ten University of California campuses, campus libraries invest more than \$60 million each year licensing e-resources; that aggregate sum represents license agreements made at the systemwide and multi-campus levels, but because each individual campus also licenses electronic resources, actual expenditures on e-resource licensing may be increased by several million dollars more.<sup>57</sup> UC Berkeley alone spends more than \$12 million.<sup>58</sup>

Because e-resource agreements are such big business for publishers—yielding profit margins for the largest academic publishers like Elsevier of around 40%, which rival that of Google and Apple<sup>59</sup>—the e-resource agreements are highly complex, and often fastidiously drafted and negotiated by publishers’ global or regional counsel. By

<sup>54</sup> Statistics on file with author.

<sup>55</sup> See the discussion of Fair Uses like TDM being the subject of additional costs in Section II. B.

<sup>56</sup> ASSOCIATION OF RESEARCH LIBRARIES, STATISTICS (2024) (on file with author). Note that not all “ongoing” subscriptions are electronic resources, but the majority are. Figures from 2021 for U.S. and Canadian research libraries that are ARL members tallied \$1.7 billion annually, showing continued expansion and growth. *ARL Board Affirms Transparency as Core Operating Principle of Licensing*, ASS’N OF RESEARCH LIBRARY. (Aug. 23, 2021),

<https://www.arl.org/news/arl-board-affirms-transparency-as-core-operating-principle-of-licensing/>

<sup>57</sup> Rachael Samberg, Timothy Vollmer & Samantha Teremi, *Fair Use Rights to Conduct Text and Data Mining and Use Artificial Intelligence Tools are Essential for UC Research and Teaching*, U.C. OFF. OF SCHOLARLY COMM’N. BLOG (Mar. 12, 2024),

<https://osc.universityofcalifornia.edu/2024/03/fair-use-tdm-ai-restrictive-agreements/>. The University of California campuses have different research focus areas and different budgets, so not all resources are licensed by all ten campuses. As a result, campuses like UC Berkeley can invest millions of dollars more licensing content specifically requested by UC Berkeley scholars.

<sup>58</sup> *Id.*

<sup>59</sup> See Kayla Yup, *How Scientific Publishers’ Extreme Fees Put Profit Over Progress*, THE NATION (Aug. 5, 2024),

<https://www.thenation.com/article/society/neuroimage-elsevier-editorial-board-journal-profit/> (Elsevier publishing profits near 40%, rivaling Apple and Google); SPARC, LANDSCAPE ANALYSIS REP. (Mar. 29, 2019), <https://infrastructure.sparcopen.org/landscape-analysis/elsevier> (Elsevier profits around 40%; Springer-Nature around 23%). See also Martin Hagve, *The Money Behind Academic Publishing*, TIDDSKRIFTET (Aug. 17, 2020), <https://tidsskriftet.no/en/2020/08/kronikk/money-behind-academic-publishing>.



contrast, most academic institutions cannot afford to employ attorneys to negotiate the hundreds of e-resource agreements they sign each year, and instead rely on librarians to secure essential research rights.<sup>60</sup> There has been some growth in the presence of lawyers-cum-librarians within academic libraries who also bear copyright and contracts negotiation expertise and can oversee this work.<sup>61</sup> There has also been some expansion in the availability of educational resources to support better library licensing outcomes across institutions.<sup>62</sup> By-and-large, however, publishers and their attorneys have far greater power and expertise in the contracting process than academic research libraries.<sup>63</sup> Libraries are thus often at a disadvantage in endeavoring to ensure they adequately address a myriad of copyright-dependent needs for their campus scholars.<sup>64</sup>

Infrequently, libraries can negotiate a broad set of licensed rights (in a license “grant” clause) that inherently covers all of the above uses without any further qualification. The California Digital Library—the entity that licenses certain content

<sup>60</sup> The responsibility of licensing falling to librarians who are not legal experts is addressed in SANDRA ENMIL, RACHAEL SAMBERG, SAMANTHA TEREMI, KATIE ZIMMERMAN & ERIK LIMPITLAW, *E-RESOURCE LICENSING EXPLAINED* (2024),

<https://berkeley.pressbooks.pub/eresourcelicensingexplained>. The general licensing responsibilities of non-legal-expert librarians are also set forth in books like: COREY S. HALAYCHIK & BLAKE REAGEN, *LICENSING ELECTRONIC RESOURCES IN ACADEMIC LIBRARIES: A PRACTICAL HANDBOOK* (2018); ALBITZ, *supra* note 53, at 51 (2008); HARRIS, *supra* note 53.

<sup>61</sup> UNIVERSITY INFORMATION POLICY OFFICERS, <https://uipo.org/> (last visited June 5, 2025) [hereinafter UIPO] is an organization for copyright and information policy professionals working in academic libraries and research libraries in the United States and Canada. Members are from nearly 40 states and provinces, working in over 75 institutions and organizations.”) *See also The Journal of Copyright in Education and Librarianship*, Vol. 6, No. 1 (2022) (detailing pathways to copyright librarianship for lawyers and legal experts).

<sup>62</sup> *See, e.g.*, SANDRA ENMIL, RACHAEL SAMBERG, SAMANTHA TEREMI, KATIE ZIMMERMAN & ERIK LIMPITLAW, *E-RESOURCE LICENSING EXPLAINED* (2024), [hereinafter Enmil, et. al.] <https://berkeley.pressbooks.pub/eresourcelicensingexplained>.

<sup>63</sup> COMMUNIA, *supra* note 14, at 14 (“due to the power imbalance between publishers and libraries, publishers often include language to restrict uses that are permitted under the laws of the library’s home country, knowing that libraries often lack the bargaining power to properly negotiate changes to such clauses”); Bowers, *supra* note 10, at 79 (“Today’s academic libraries are forced to enter into a significant number of licensing agreements every year for access to electronic resources. Vendors, or licensors, have the power to dictate the terms and conditions set forth in the license agreement. While libraries may attempt to negotiate and change certain terms, their attempts are often futile or met with strong resistance.”); Svetlana Yakovleva, *Literature Review on the Use of Licenses in Library Context, and the Limitations This Creates to Access to Knowledge* (Jan. 20, 2017), <https://ssrn.com/abstract=3305972> (“A number of legal scholars have observed that in licensing relationships with content providers of copyright and related rights material, libraries are often the weaker party. As a result, ‘[i]t is not uncommon for right holders to wield their bargaining power to arrive at contractual terms that purport to set aside the privileges that the law grants users pursuant to the limitations on copyright’” (quoting Lucie Guibault, *Why Cherry-picking Never Leads to Harmonisation: the Case of the Limitations on Copyright under Directive 2001/29/EC*, 1 JIPITEC 55 (2010))

<sup>64</sup> *Id.*

systemwide for participating University of California campuses—developed this “model” clause for a broad grant of usage rights<sup>65</sup>:

Licensee and Authorized Users may make all use of the Licensed Materials as is consistent with United States copyright law, including its Fair Use Provisions.<sup>66</sup>

Alternatively, libraries might achieve broad rights through “fair use savings” clauses—clauses that endeavor to preserve the rights users have under Section 107 by clarifying that other contractual provisions should not be interpreted to restrict Fair Use. A typical “fair use savings clause” that a library might seek is:

Nothing in this Agreement shall be interpreted to diminish the rights and privileges of the Licensee or Authorized Users with respect to any of the Licensed Materials, including exceptions or limitations to the exclusive rights of copyright owners under the United States Copyright Act of 1976, as amended (17 U.S.C. Sec. 101 et seq.), such as fair use.<sup>67</sup>

Almost invariably, however, publishers insist on following up any such “license grant” or “fair use savings” clause by individually enumerating both “authorized uses” and “prohibited uses” in separate follow-on clauses, which control over a general fair use rights reservation.<sup>68</sup> These follow-on clauses parse uses granularly, distinguishing between: interlibrary loan, printing, caching, display, classroom uses, print course packs, electronic course packs, electronic links, scholarly sharing, bibliographic citations, and more.<sup>69</sup> Indeed, virtually all e-resource agreements signed by California Digital Library

<sup>65</sup> Model License Agreement, CAL. DIG. LIBRARY., [https://cdlib.org/wp-content/uploads/2017/01/CDL\\_Model\\_License\\_2016\\_public\\_version\\_final.docx](https://cdlib.org/wp-content/uploads/2017/01/CDL_Model_License_2016_public_version_final.docx).

<sup>66</sup> Katie Zimmerman, *Fair Use Savings Clauses*, in E-RESOURCE LICENSING EXPLAINED (2024), <https://berkeley.pressbooks.pub/eresourcelicensingexplained/chapter/fair-use-savings-clauses/>.

<sup>67</sup> *Id.*

<sup>68</sup> ALBITZ, *supra* note 53, at 19 (2008) (“Rarely, however, will an information provider agree to the inclusion of such a broad-based clause. Rather, it will want to negotiate each use individually.”). For the principle that more specific language in an agreement controls over general language on a given point, *see, e.g.* Continental Cas. Co. v. Zurich Ins. Co., 370 P.2d 198 (Cal. 1961); Bowmer v. Bowmer, 50 N.Y.2d 288 (1980); Restatement (Second) of Contracts § 203(c) (1981) (“specific terms and exact terms are given greater weight than general language”); 11 WILLISTON ON CONTRACTS § 32:10 (4th ed. 2025) (accord). For a broader discussion of the limitations of fair use savings clauses, *see* Katie Zimmerman, *Fair Use Savings Clauses*, in E-RESOURCE LICENSING EXPLAINED (2024).

<sup>69</sup> *See* License Agreement Between the Regents of the University of California and American Chemical Society, [https://cdlib.org/services-groups/collections/licensed\\_resources/redacted\\_licenses/STACS\\_Access\\_Agrmt\\_w\\_TDM\\_Rider\\_2016\\_Redacted.pdf](https://cdlib.org/services-groups/collections/licensed_resources/redacted_licenses/STACS_Access_Agrmt_w_TDM_Rider_2016_Redacted.pdf) (“Permitted Uses” vs. “Prohibited Uses”); License Agreement Between the Regents of the University of California and Wiley, <https://ucsf.app.box.com/s/in1d6ry2vc2tryvypb7yz3ofg90zhy6q> (“Authorized Uses” vs. “Prohibited uses”); License Agreement Between the Regents of the University of California and Springer Nature, <https://ucop.app.box.com/s/v1260wmpe8pmm1jepwj36t7u1k1cssi8> (within

on behalf of the University of California libraries reflect such delineations after the general license grant clause.<sup>70</sup>

Given the many ways that scholars make use of copyright-protected works, publisher nullification of copyright exceptions can significantly impact the work that scholars do. For example, one of Liverpool University Press' agreements with UC Berkeley governing e-books precludes interlibrary loan even of individual chapters: "Neither the Licensee nor Authorised nor Walk-in Users may provide, by electronic means, to a user at another institution or elsewhere, a retained electronic copy **of any part** of the Licensed Material" (emphasis added).<sup>71</sup> The circumvention of interlibrary loan in this manner can have a profound effect on scholarly work:

While the library may have retained its fair use rights which would allow for a copy to be made in certain situations, the lack of rights to engage in interlibrary loan preempts the ability of anyone who is not an authorized user under the license agreement to access the electronic content. As a result, the tradition of libraries freely lending, borrowing, and exchanging portions of their collections with other libraries is restricted or eliminated... When the library is not able to engage in interlibrary loan transactions as a result of the license agreement, many of its potential patrons endure a great disservice - severely limited access. This license restriction results in the curtailment of one of the foundations of libraries, interlibrary loan.<sup>72</sup>

S&P Global, a financial data and analytics company that provides essential content for business and social sciences researchers, prohibits interlibrary loan and the sharing of even portions of any content with any scholarly collaborators, yet most remarkably also prohibits the publishing any portions of content, including by the licensee:

The Service(s) is solely and exclusively for the non-commercial and/or academic use of Authorized Users. Access to the Service(s) or any portion thereof shall be limited to those Authorized Units identified on the Exhibit A. Further, except as expressly permitted herein or in an Exhibit A, Subscriber shall not (a) license, sublicense, transfer, sell, resell, **publish, reproduce, and/or otherwise redistribute the Services or any components thereof in any manner** (including but not limited to via or as part of any

---

General Terms attachment, "Grant of License" vs. "Prohibited Uses"); License Agreement Between the Regents of the University of California and FACTIVA, [https://cdlib.org/services-groups/collections/licensed\\_resources/redacted\\_licenses/ST\\_Tier2\\_Factiva\\_UCLA\\_2005\\_Redacted.pdf](https://cdlib.org/services-groups/collections/licensed_resources/redacted_licenses/ST_Tier2_Factiva_UCLA_2005_Redacted.pdf) ("Permitted Users May" vs. "Permitted Users May Not"); *see also supra* note 70.

<sup>70</sup> *Id.*

<sup>71</sup> *See* Liverpool University Press, Site License Agreement (2023 Translated Texts for Historians E-Library), <https://drive.google.com/file/d/1aHDcjo0U-KcFrGTrijmTty2nom8VfEK/view?usp=sharing>. *See also* African Books Collective, Ebook Distribution End User Agreement (July 2025), [https://drive.google.com/file/d/1FPuDGqVj23ta1342JN\\_SGqDMZAAy5uLe/view?usp=drive\\_link](https://drive.google.com/file/d/1FPuDGqVj23ta1342JN_SGqDMZAAy5uLe/view?usp=drive_link) ("5.4 Interlibrary Loan. It is not permissible for the Licensed Materials to be made available for Interlibrary Loan.")

<sup>72</sup> *See, e.g., Bowers, supra* note 10, at 84.

intranet site) or (b) **provide access to the Services or any portion thereof to any person, firm or entity** other than the Authorized Unit<sup>73</sup>

CRSP (Center for Research in Security Prices), when licensing integral business content for academic research, similarly prohibits any publishing or reproduction:

For the purposes of this Agreement, “Internal Use” means use for the Subscriber’s benefit by faculty, students, and staff at the specific campus identified in the preamble above, and on one computer system or approved server and group of networked workstations in conjunction with software applications **solely for the purposes of researching and analyzing stock trends, including but not limited to, preparation of reports, but expressly excludes further dissemination of the information or data** contained in the CRSP® Data Files in electronic form (except as research results derived from the CRSP® Data Files... **Subscriber agrees that it will not transfer, sell, publish, redistribute or release or otherwise make available the CRSP® Data Files or the data contained therein to any individual or third party who is not a of Subscriber without having received express written permission from CRSP® in advance.**<sup>74</sup> (emphasis added)

The standard academic license agreement for African Book Collective, a collective “owned by its founder publishers” and serving as a “worldwide marketing and distribution outlet” for scholarly books from Africa,<sup>75</sup> precludes scholars from copying or pasting any content from the books, except as determined by the publisher: “The extent to which the Licensee and Authorised Users shall be allowed to Print, Copy and Paste is determined by the Licensor.”<sup>76</sup> In addition, no portions or snippets of the licensed works may be downloaded or published even for research purposes.<sup>77</sup>

<sup>73</sup> See S&P Global Market Intelligence, Academic Subscription Agreement, <https://drive.google.com/file/d/1YUanRuUe85itcB2BpUezhi3KLXZEYWh/view?usp=sharing>.

<sup>74</sup> See CRSP, Data Standard Subscription Agreement, [https://drive.google.com/file/d/1hbC54Ls6lOOt2jb1c\\_byp022ETHdL3Du/view?usp=sharing](https://drive.google.com/file/d/1hbC54Ls6lOOt2jb1c_byp022ETHdL3Du/view?usp=sharing) (Pars. 1.2 and 1.3; see also Appendix D: CRSP Survivor-Bias-Free US Mutual Fund Database, providing that “Display, performance, reproduction, distribution of, or creation of derivative works or improvements from the Thomson Reuters data contained in CRSP®’s Survivor-Bias-Free US Mutual Fund Database in any form or manner is expressly prohibited, except to the extent expressly permitted hereunder, or otherwise, with the prior written permission of Thomson Reuters...Subscriber may copy, paste and distribute only an insubstantial amount of the data contained in the CRSP®’s Survivor-Bias-Free US Mutual Fund Database provided that: **(b) the data is not distributed by Subscriber in connection with information vending or commercial publishing (in any manner or format whatsoever), nor reproduced through the press or mass media or on the Internet**”) (emphasis added).

<sup>75</sup> About, AFRICAN BOOK COLLECTIVE, <https://africanbookscollective.com/about/> (last visited July 16, 2025).

<sup>76</sup> African Books Collective, Ebook Distribution End User Agreement (July 2025), [https://drive.google.com/file/d/1FPuDGqVj23taI342JN\\_SGqDMZAAY5uLe/view?usp=drive\\_link](https://drive.google.com/file/d/1FPuDGqVj23taI342JN_SGqDMZAAY5uLe/view?usp=drive_link) (Par. 5.5)

<sup>77</sup> *Id.* (Par. 3.4.7: The Licensee and Authorized Users may not “...download, mount or distribute any part of the Licensed Material on any electronic system or network.”; Par. 3.4.9: The Licensee and Authorized Users may not “Publish, distribute or make available the Licensed Materials,

To be clear: Under the terms of each of these agreements, the publishers or vendors are withdrawing even the ability to quote from or include excerpts from the licensed content in published research—acts which are quintessential fair uses.

While the impact of preventing interlibrary loan or publishing excerpts may be relatively self-evident, perhaps the best way to illustrate the far-reaching impact of publishers' curtailment of fundamental rights is by profiling how e-resource agreements endeavor to abrogate fair use in the specific context of what is referred to as "computational research" methodologies (also called "text and data mining" or "TDM")—and especially TDM that relies on artificial intelligence (AI).

### *B. TDM and AI clauses that override fair use*

With increasing importance and frequency, scholars have employed computational tools, algorithms, and automated techniques to extract revelatory information from large sets of unstructured or thinly-structured digital content.<sup>78</sup> This process is broadly known as TDM, and it allows researchers to identify and analyze patterns, trends, and relationships across volumes of data that would otherwise be impossible to sift through.<sup>79</sup> TDM enables the exploration of issues like: racial disparity evidenced through police body camera footage;<sup>80</sup> changes in gender significance in fiction;<sup>81</sup> and public discussions of social justice issues like violence against women.<sup>82</sup> The particular TDM methodologies employed continue to expand, enabling advancements across education, literature, society, politics, and beyond.<sup>83</sup>

Not all TDM research methodologies necessitate usage of AI models. For instance, sometimes TDM can be performed by developing algorithms to detect the frequency of certain words within a corpus, or to parse sentiments based on the proximity of various words to each other.<sup>84</sup> In other cases, though, scholars must employ machine learning techniques to train AI models before the models can make a variety of assessments.

---

works based on the Licensed Materials or works which combine them with any other material... without the Licensors' Express Permission.")

<sup>78</sup> Peter McCracken & Ellen Raub, *Licensing Challenges Associated With Text and Data Mining: How Do We Get Our Patrons What They Need?*, 11 J. LIBRARY. SCHOLARLY COMM'N., no. 1, 2023, at 2.

<sup>79</sup> Marti A. Hearst, *What is text mining?*, U.C., BERKELEY I-SCHOOL (Oct. 17, 2003), <http://people.ischool.berkeley.edu/~hearst/text-mining.html>.

<sup>80</sup> Rob Voigt, et al., *Language From Police Body Camera Footage Shows Racial Disparities in Officer Respect*, 114 PROCEEDINGS NAT'L ACAD. SCI. 6521 (2017).

<sup>81</sup> Ted Underwood, David Bamman & Sabrina Lee, *The Transformation of Gender in English-Language Fiction*, CULTURAL ANALYTICS, Feb. 13, 2018, at 1.

<sup>82</sup> Jia Xue et al., *Harnessing Big Data for Social Justice: An Exploration of Violence Against Women-Related Conversations on Twitter*, HUMAN BEHAVIOR AND EMERGING TECHNOLOGIES, 1 HUM. BEHAV. EMERGING TECH. 269 (2019).

<sup>83</sup> Hossein Hassani et al. *Text Mining in Big Data Analytics*. 4 BIG DATA COGNITIVE COMPUT., no. 1, 2020, at 1.

<sup>84</sup> See, e.g., *Google Books nGram Viewer*, GOOGLE, <https://books.google.com/ngrams/info> (last visited June 6, 2025); *Sentiment Analysis*, WIKIPEDIA.COM, [https://en.wikipedia.org/wiki/Sentiment\\_analysis](https://en.wikipedia.org/wiki/Sentiment_analysis) (last visited June 6, 2025).

These AI models may be either non-generative or generative: the former are reliant only on discriminative modeling (classification and regression) while the latter systems' generative modeling is capable of producing new outputs such as text or images.<sup>85</sup> According to the U.S. Copyright Office, non-generative AI systems "have already produced miracles in scientific and medical research" and many generative AI applications promise similarly great benefits for the public.<sup>86</sup>

The following example illustrates the distinction between "regular" TDM and "AI-infused" TDM research: Imagine a scholar wishes to assess the prevalence with which 20th century films mention gender or related words. This scholar likely would compile a corpus of thousands or tens of thousands of films, and then run a search algorithm across the corpus to detect the occurrence or frequency of the word "gender" and its synonyms. But if the scholar instead wanted to establish the *presence* of individuals representing a certain gender in films, the scholar would first need to train an AI model to recognize the appearance of various genders in a smaller corpus of films and then, once trained, ask the model to compute the presence of those genders in a much broader body of works.<sup>87</sup> In this way, AI modeling can be used to investigate the representation of race, gender, and more in both popular and prestige films and TV shows, and answer questions such as: How are race and gender tied to the depiction of characters on screen, and how has this changed over the past 50 years? Indeed, this work was recently undertaken in reliance on a new regulation that for the first time enables bypassing digital rights management on DVDs for scholarly research.<sup>88</sup>

As an exception to copyright's grant of exclusive rights to copyright owners, fair use is particularly essential for TDM and "AI-infused" TDM for the following reasons: If a scholar could not rely on fair use to make the reproductions necessary for computational research, then such research would be permitted only on "safe" materials (like public domain works, or works for which training permission has been granted via a license). Studies have shown that this would curtail freedom of inquiry, exacerbate bias in the nature of research questions able to be studied and the methodologies available to study them, and amplify the views of an unrepresentative set of creators given the limited types of materials available with which to conduct the studies.<sup>89</sup> And this is a particular problem for TDM because TDM research often requires "the use of "massive datasets with works from many publishers, including copyright owners that

<sup>85</sup> Katherine Lee, A. Feder Cooper & James Grimmelman. *Talkin' 'Bout AI Generation: Copyright and the Generative AI Supply Chain*, J. COPYRIGHT SOC. (forthcoming), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4523551](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4523551) [<https://doi.org/10.2139/ssrn.4523551>].

<sup>86</sup> U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE PART 3: GENERATIVE AI TRAINING (pre-publication version May 2025), fn. 413 (citing e.g., *Stopping Malaria in Its Tracks*, GOOGLE DEEPMIND (Oct. 13, 2022), <https://deepmind.google/discover/blog/stopping-malaria-in-its-tracks/>).

<sup>87</sup> For an understanding of discriminative modeling (classification and regression) vs. generative modeling (systems capable of producing outputs such as text or images), see Katherine Lee, A. Feder Cooper & James Grimmelman. *Talkin' 'Bout AI Generation: Copyright and the Generative AI Supply Chain*, J. COPYRIGHT SOC. U.S.A. (forthcoming).

<sup>88</sup> 37 C.F.R. § 201.40.

<sup>89</sup> Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579 (2018).

cannot be identified or are unwilling to grant licenses.”<sup>90</sup> Very little content would be available to study.

Text and data mining in these contexts is clearly a fair use. Several court cases, including *Authors Guild v. HathiTrust* and *Authors Guild v. Google*, have confirmed this over the last two decades,<sup>91</sup> and the U.S. Copyright Office has agreed.<sup>92</sup> This is in large part due to the transformativeness of the purpose.<sup>93</sup> As these cases note, when the use is highly transformative under Factor One—as courts have found corpus creation to extract patterns and information to be—“the more it serves copyright’s goal of enriching public knowledge and the less likely it is that the appropriation will serve as a substitute for the original or its plausible derivatives” for purposes of Factor Four.<sup>94</sup> This is even true where, as in those cases, the subject use involved copying *millions* of copyrighted books.<sup>95</sup> Further, the cases have held that distributing derived data, results, abstractions, metadata, or analysis from the corpus is also fair use, as long as the research methodologies or data distribution processes do not re-express the underlying copyrighted works to the public in a way that could supplant the market for the originals.<sup>96</sup>

For the same reasons that TDM processes constitute fair use of copyrighted works in noncommercial research contexts, the Copyright Office has recently opined that

<sup>90</sup> Sean M. Fiil-Flynn et al., *Legal Reform to Enhance Global Text and Data Mining Research*, 378 SCIENCE 951 (2022) (internal citations omitted).

<sup>91</sup> See *Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014) (creation of a full-text searchable digital library that displays word count results), *affirming* 902 F. Supp. 2d 445, 460 (S.D.N.Y. 2012) (“The use to which the works in the HDL are put is transformative because the copies serve an entirely different purpose than the original works: the purpose is superior search capabilities rather than actual access to copyrighted material. The search capabilities of the HDL have already given rise to new methods of academic inquiry such as text mining.”); *Authors Guild v. Google*, 804 F.3d 202 (2d Cir. 2015) (creation of a full-text searchable database with “snippet view” and “ngram viewer” [search strings]); *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630 (4th Cir. 2009) (use of machine learning on a corpus so that the AI can detect the occurrence of plagiarism in other inputs).

<sup>92</sup> In evaluating a proposed DMCA § 1201 exemption to circumvent technological protection measures on DVDs and e-books for the purpose of conducting TDM—which exemption the Copyright Office has granted and renewed—the Copyright Office writes: “Balancing the four fair use factors, with the limitations discussed, the Register concludes that the proposed use is likely to be a fair use.” See U.S. COPYRIGHT OFFICE, SECTION 1201 RULEMAKING: EIGHTH TRIENNIAL PROCEEDING TO DETERMINE EXEMPTIONS TO THE PROHIBITION ON CIRCUMVENTION - RECOMMENDATION OF THE REGISTER OF COPYRIGHTS (Oct. 2021), at 117.

<sup>93</sup> 755 F.3d at 97 (“[T]he creation of a full-text searchable database is a quintessentially transformative use. [T]he result of a word search is different in purpose, character, expression, meaning, and message from the page (and the book) from which it is drawn. Indeed, we can discern little or no resemblance between the original text and the results of the HDL full-text search... Full-text search adds a great deal more to the copyrighted works at issue than did the transformative uses we approved in several other cases”); 804 F.3d 216-219 (accord).

<sup>94</sup> *Authors Guild v. Google*, 804 F.3d at 214.

<sup>95</sup> 804 F.3d at 207 (“Google has made digital copies of tens of millions of books”); 755 F.3d 87 (more than 10 million digitized works at issue).

<sup>96</sup> 804 F.3d at 226 (lawful to “allow the public to obtain limited data about the contents of the book, without allowing any substantial reading of its text”).



training AI tools to conduct that TDM for nonprofit research and educational uses is also fair use.<sup>97</sup> There is reason to believe that a court would follow this guidance in similar circumstances. In *A.V. ex rel. Vanderhye v. iParadigms*,<sup>98</sup> the court addressed text mining as part of plagiarism detection software for schools.<sup>99</sup> The software at issue relied on machine learning—a form of AI—to be able to undertake that plagiarism detection.<sup>100</sup> The court ultimately found that computational analysis involving machine learning (and thus AI) as one of its component steps was fair use.<sup>101,102</sup>

<sup>97</sup> U.S. COPYRIGHT OFFICE, COPYRIGHT AND ARTIFICIAL INTELLIGENCE PART 3: GENERATIVE AI TRAINING (pre-publication version May 2025), at 46, 74 (re: Factor One: “On one end of the spectrum, training a model is most transformative when the purpose is to deploy it for research, or in a closed system that constrains it to a non-substitutive task. For example, training a language model on a large collection of data, including social media posts, articles, and books, for deployment in systems used for content moderation does not have the same educational purpose as those papers and books.”; re: Fair Use overall: “On one end of the spectrum, uses for purposes of noncommercial research or analysis that do not enable portions of the works to be reproduced in the outputs are likely to be fair.”)

<sup>98</sup> 562 F.3d 630 (4th Cir. 2009).

<sup>99</sup> *Id.* at 634.

<sup>100</sup> Although the term “machine learning” is not mentioned in the opinion, scholars understand the iParadigms algorithm to function only having once been trained to characterize passages.

<sup>101</sup> 562 F.3d at 645.

<sup>102</sup> We take no position on whether the same result should or would be true regarding AI training in the commercial context, as those uses are not at issue for the libraries and scholars discussed herein. Nevertheless, we note that district courts to date have found that AI training for commercial LLM tools can also be fair when the intent is not to create a competing commercial product. In *Bartz v. Anthropic*, No. C 24-05417 WHA (N.D. Cal. June 23, 2025), Anthropic digitized plaintiff’s copyrighted works (millions of books) to train its large language model. Plaintiff argued, “and this order takes for granted — that such training entailed ‘memoriz[ing]’ works by ‘compress[ing]’ copies of those works into the LLM... Regardless, the ‘purpose and character’ of using works to train LLMs was transformative — **spectacularly so.**” *Id.* at 11 (emphasis added). Overall, use was fair because of transformativeness under factor one and no market supplantation under factor four: **For factor one**, “Anthropic’s LLMs have not reproduced to the public a given work’s creative elements, nor even one author’s identifiable expressive style (assuming arguendo that these are even copyrightable)...If this training process reasonably required making copies within the LLM or otherwise, those copies were engaged in a transformative use.” *Id.* at 12-14. **For factor four**, “The copies used to train specific LLMs did not and will not displace demand for copies of Authors’ works, or not in the way that counts under the Copyright Act...Authors next contend that training LLMs displaced (or will) an emerging market for licensing their works for the narrow purpose of training LLMs...A market could develop...Even so, such a market for that use is not one the Copyright Act entitles Authors to exploit.” *Id.* at 28. Likewise, in *Kadrey v. Meta Platforms*, another district court considered AI training in the commercial context (involving the use of “pirated” books online), found that AI training in this context was transformative under factor one, though, had evidence of damages based on a theory of “market dilution” been introduced, theoretically could have caused market disruption under factor four; yet even as to factor four and the “market dilution” approach, the court distinguished the research context as not having the same market effect. *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417-VC at 36, fn14 (N.D. Cal. June 26, 2025) (order granting defendant’s motion for summary judgment); *Cf. Thomson Reuters Enter. Ctr. GmbH v. Ross Intel.*

The precedents set by *Authors Guild v. Google* and *Authors Guild v. Hathitrust* should ensure that scholars within the U.S. can undertake the copying and reuse of materials necessary for TDM. Further, because these cases involved scanning copies of print version books for TDM, they did not involve problematic contracts overriding those underlying rights afforded by copyright exceptions. But that is also the problem: libraries cannot simply scan all the print version books and all the print version journal articles in the world that their researchers want to use for TDM.<sup>103</sup> Not only is it technologically impossible, but it is financially infeasible.<sup>104</sup> Google invested millions of dollars to digitize physical books<sup>105</sup>, and there are now millions more books, and tens of millions of scholarly journal articles since that project began.<sup>106</sup> While libraries continue to acquire physical books and continue to scan them for TDM purposes, libraries do not have the extraordinary resources that would be required to purchase print versions of everything and subsequently undertake digitization according to *Authors Guild v. Google* and *Authors Guild v. Hathitrust* parameters. If libraries and the researchers they serve want digital access to content, often the only possible pathway is to obtain a license from the publisher.

Unfortunately, these licensing pathways are often foreclosed for TDM and AI training in research. Fair uses are becoming an out-of-reach licensing luxury even for the wealthiest institutions, as increasingly academic libraries are forced to pay significant sums each year to try to preserve fair use rights for campus scholars within

---

Inc., 765 F. Supp. 3d 382 (training a commercial AI tool to create a competing product was not Fair Use).

<sup>103</sup> See, e.g., Authors Alliance, Text and Data Mining under U.S. Copyright Law: Landscape, Flaws & Recommendations, at 9-10 (Oct. 2024), <https://www.authorsalliance.org/wp-content/uploads/2024/11/Text-and-Data-Mining-Report-102024.pdf> (costs of digitization and optical character recognition along with metadata generation are often prohibitive).

<sup>104</sup> *Id.*

<sup>105</sup> In 2006, the estimate for what Google would invest to digitize 30 million books was \$750,000,000. See Jonathan Band, *The Google Library Project: Both Sides of the Story*, PLAGIARY, 6-21 (2006).

<sup>106</sup> By 2019, Google had digitized 40 million books. 15 Years of Google Books, GOOGLE (Oct. 17, 2019), <https://blog.google/products/search/15-years-google-books/>. An estimate from 2010 put the number of scholarly journal articles at 50 million. A.E. Jinha, *Article 50 Million: An Estimate of the Number of Scholarly Articles in Existence*, 23 LEARNED PUBL'G 258 (2010). More than fifteen years have elapsed since that calculation.

the content license agreements that libraries sign, and many are unsuccessful.<sup>107</sup> When such costs are beyond institutional reach, the publisher or vendor may then offer similar contractual terms directly to research teams, who may feel obliged to agree in order to get access to the content they need.<sup>108</sup> Vendors may charge tens or even hundreds of thousands of dollars for this type of access.<sup>109</sup> Given the sheer volume of licensed content needed for TDM and AI-infused TDM, this “pay-to-play” landscape of charging institutions for the opportunity to rely on existing statutory rights is particularly detrimental for TDM research methodologies.<sup>110</sup>

To better understand the impact of such prohibitions on underlying rights, it may be helpful to decipher examples of restrictive licensing language presented to academic libraries as part of these agreements. On one hand, some publishers’ provisions simply prohibit TDM or the use of AI full-stop. For instance, the University of California, Berkeley (UCB) recently received the following proposed language from one of the world’s leading scholarly content aggregators, Clarivate:

Except as expressly permitted in this Addendum you [i.e. UCB] and your Authorized Users [i.e. UCB faculty, staff, students] shall not: i) Text mine, data mine or harvest metadata from the Product, use the Product or underlying data in conjunction with any third-party technology or any artificial intelligence, algorithms or models, or use the

<sup>107</sup> In a white paper on copyright and contractual issues faced by libraries, the Association of Research Libraries states, “Arguably, contract terms that seek to limit exceptions under the Copyright Act

are preempted under a conflict-preemption theory” but notes there is “surprisingly little case law on this point.” Katherine Klosek, *Copyright and Contracts: Issues & Strategies*, ASS’N OF RSCH. LIBRARY. (2022),

<https://www.arl.org/wp-content/uploads/2022/07/Copyright-and-Contracts-Paper.pdf>. By contrast, this is not the case in the European Union. As discussed in Section IV.A.i., *infra*, the EU’s Copyright in the Digital Single Market Directive provides that, “Any contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 shall be unenforceable,” referencing articles that govern text mining and data mining, digital cross-border teaching, and preservation by cultural heritage organizations, respectively. *See* Council Directive 2019/790, 2019 O.J. (L 130) 92 (EU); *see also* Singapore’s Copyright Bill, which similarly prohibits contractual override of copyright exceptions. *Copyright Bill*, 2021, Bill 17 (Sing.).

<sup>108</sup> Klosek, *supra* note 108; McCracken & Raub, *supra* note 79, at 8 (“Gale, ProQuest, JSTOR, and LexisNexis have each created their own online, fee-based platforms, which they hope libraries will use for most of their TDM work. Each space is a revenue generating platform in which patrons will learn how to perform text analysis and also a space through which the vendor can limit and control access to its proprietary content.”)

<sup>109</sup> *Id.* *See also* Katherine Klosek & Samantha Teremi, *AI Is Reigniting Decades-Old Questions Over Digital Rights, but Fair Use Prevails*, ASS’N OF RSCH. LIBRARY. (Feb. 28, 2025), <https://www.arl.org/blog/ai-is-reigniting-decades-old-questions-over-digital-rights-but-fair-use-prevails/> (“A publisher recently provided UC Berkeley’s Library with an elusive explanation for their AI ban on a subset of their licensed materials, claiming that they would ‘require **new and different AI terms [that] would be significantly higher in price.**’ (emphasis added)”)

<sup>110</sup> Sean M. Fiil-Flynn et al., *Legal Reform to Enhance Global Text and Data Mining Research*, 378 SCIENCE 951 (2022) (internal citations omitted).

Product or underlying data to develop or train any artificial intelligence, algorithms or models.<sup>111</sup>

That provision would ban all TDM, as well as any use or training of any AI tool that either UCB scholars create, or any AI tool created by any other scholar. At the same time, Clarivate licenses and distributes scholarship that relies on these very methodologies—meaning at some level they recognize the value of these fair use methodologies like TDM and AI training for knowledge creation.<sup>112</sup>

Not all TDM and AI fair use overriding provisions are as obvious in their effect as the Clarivate prohibition. For example, leading scholarly publisher Springer-Nature had originally proposed the following to the University of California:

Customer [the university] and its Authorized Users [the scholars] may not:

1. directly or indirectly develop, train, program, improve, and/or enrich any artificial intelligence tool (“AI Tool”) accessible to anyone other than Customer and its Authorized Users, whether developed internally or provided by a third party; or
2. reproduce or redistribute the Content to any third-party AI Tool, except to the extent limited portions of the Content are used solely for research and academic purposes (including to train an algorithm) and where the third-party AI Tool (a) is used locally in a self-hosted environment or closed hosted environment solely for use by Customer or Authorized Users; (b) is not trained or fine-tuned using the Content or any part thereof; and (c) does not share the Content or any part thereof with a third party.

This Springer-Nature language might initially seem preferable to Clarivate’s flat ban, but the consequences (had the University of California not successfully negotiated

<sup>111</sup> See Clarivate, Terms Master Client Agreement & Product Service Terms Addendum, [https://drive.google.com/file/d/152pS-H9NHsfPw9xtdMkIbF947lGUvXLT/view?usp=drive\\_link](https://drive.google.com/file/d/152pS-H9NHsfPw9xtdMkIbF947lGUvXLT/view?usp=drive_link) (Par. 2(g); Product/Services Terms Addendum Pars. 4 & 12). Many other agreements assert the same AI prohibitions. See, e.g., S&P Global Market Intelligence, Academic Subscription Agreement, <https://drive.google.com/file/d/1YUanRuUe85itcB2BpUezhi3KLXZEvYWh/view?usp=sharing> (Par. 1.4: “Subscriber shall not use the Services, any derivatives thereof, or any part thereof in any software or systems, including artificial intelligence models and systems, that: (a) exposes the same to third parties.”) For additional sample clauses in such agreements, see, e.g., COMMUNIA, *supra* note 14.

<sup>112</sup> Laura Marrlett, *Uncovering Latent Structure and Emerging Research Themes in Plastic Waste Literature Using Natural Language Processing* (Ph.D. dissertation, Arizona State University 2025), <https://www.proquest.com/dissertations-theses/uncovering-latent-structure-emerging-research/docview/3202691323/se-2?accountid=1449>; Steve Ho, *Applied Machine Learning for Big Data Genomics* (Ph.D. dissertation, University of Michigan 2025), <https://www.proquest.com/dissertations-theses/applied-machine-learning-big-data-genomics/docview/3202735308/se-2?accountid=14496>; We also note that American Chemical Society, which imposes some of the most restrictive TDM and AI usages on its own articles, publishes articles that rely on TDM and AI training using articles from other publishers. See, e.g. Lawson T. Glasby et al., *DigiMOF: A Database of Metal–Organic Framework Synthesis Information Generated via Text Mining*, 35 J. CHEM. MATERIALS 4510 (2023); Edward Kim et al., *Chemistry of Materials*, 29 J. CHEM. MATERIALS 9436 (2017).

an alternative, discussed in Section IV.B.i, *infra*) would have been comparably detrimental. The first paragraph Springer-Nature proposed forbids the training or improving of any AI tool if it is accessible or released to third parties. It further forbids the use of any computational outputs or analysis that are derived from the licensed content from being used to train any tool available to third parties. This means that the trained tools in the following fair use-reliant projects could never have been disseminated:

1. In 2017, chemists created and trained a generative AI tool on 12,000 published research papers regarding synthesis conditions for metal oxides, so that the tool could identify anticipated chemical outputs and reactions for any given set of synthesis conditions entered into the tool.<sup>113</sup> The generative tool they created is not capable of reproducing or redistributing any licensed content from the papers; it has merely learned conditions and outcomes and can predict chemical reactions based on those conditions and outcomes. And dissemination of this beneficial tool would have been prohibited under Springer-Nature's original terms identified above.
2. In 2018, researchers trained an AI tool (that they had originally created in 2014) to understand whether a character is "masculine" or "feminine" by looking at the tacit assumptions expressed in words associated with that character.<sup>114</sup> That tool can then look at other texts and identify masculine or feminine characters based on what it knows from having been trained before. Scholars can then use texts from different time periods with the tool to study representations of masculinity and femininity over time. No licensed content, no licensed or copyrighted books from a publisher can ever be released to the world by sharing the trained tool; the trained tool is merely capable of topic modeling—but Springer-Nature's language would have prohibited its dissemination.

The second paragraph of Springer-Nature's originally-proposed language is perhaps even more concerning. It provides that when using third party AI tools of any kind, a scholar can use only limited portions of the licensed content with the tools, and are prohibited from doing any training at all of third party tools even if it is a non-generative AI tool and the scholar is performing the work in a completely closed and highly secure research environment. The impact of this would have meant a complete prohibition on fair use-dependent research projects like:

1. In 2019, authors used text from millions of books published over 100 years to analyze cultural meaning.<sup>115</sup> They did this by training third-party non-generative AI word-embedding models called Word2Vec and GLoVE on multiple textual archives. The tools cannot reproduce content: when shown new text, they merely represent words as numbers, or vectors, to evaluate or predict how similar words in a given space are semantically or linguistically. The similarity of words can reveal cultural shifts in understanding of socioeconomic factors like class over time. But the publisher's above licensing terms would prohibit the training of the tools to begin with, much less the sharing of them to support further or different inquiry.

---

<sup>113</sup> Edward Kim et al., *Chemistry of Materials*, 29 J. CHEM. MATERIALS 9436 (2017).

<sup>114</sup> Ted Underwood, David Bamman & Sabrina Lee, *The Transformation of Gender in English-Language Fiction*, 3 J. CULTURAL ANALYTICS (2018).

<sup>115</sup> Austin C. Kozlowski, Matt Taddy & James A. Evans, *The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings*, 84 AM. SOC. REV. 905 (2019).

2. In 2023, scholars trained a third-party-created open-source natural language processing (NLP) tool called Chemical Data Extractor (CDE).<sup>116</sup> Among other things, CDE can be used to extract chemical information and properties identified in scholarly papers. In this case, the scholars wanted to teach CDE to parse a specific type of chemical information from scholarly journals: metal-organic frameworks, or MoFs. Anyone can now use both the trained CDE tool and the resulting database of MoF properties to ask different chemical property questions or identify additional MoF production pathways. Neither the CDE tool nor the MoF database reproduces or contains the underlying scholarly papers that the tool learned from. Yet, the tool could not be trained or disseminated under Springer Nature's initial licensing language cited above.

Indeed, there are hundreds of AI tools that scholars have trained and disseminated—tools that do not reproduce licensed content—and that scholars have created or fine-tuned to recognize faces,<sup>117</sup> decode conversations,<sup>118</sup> infer character types,<sup>119</sup> and so much more.

UC Berkeley Library has developed and released adaptable licensing language to support academic institutions in achieving better licensing outcomes through negotiations.<sup>120</sup> Yet, the onus of advocacy and negotiation to preserve statutory rights remains on academic libraries that may lack in-house expertise. Even robust sets of attorneys negotiating for large consortia—such as is the case in Europe, where academic library consortia often negotiate usage for all academic libraries in a country—struggle to preserve research rights protected by law.<sup>121</sup> Successful negotiation becomes all the more insurmountable when librarians endeavor to license from aggregators like Clarivate or EBSCO, who have themselves signed underlying agreements with publishers and copyright owners that eviscerate fair use rights—and restrict what usage rights the aggregators can license to libraries downstream.<sup>122</sup> The result is a veritable

<sup>116</sup> Lawson T. Glasby et al., *DigiMOF: A Database of Metal–Organic Framework Synthesis Information Generated via Text Mining*, 35 J. CHEM. MATERIALS 4510 (2023).

<sup>117</sup> Wei Wu, Hao Peng & Shiqi Yu, *YuNet: A Tiny Millisecond-level Face Detector*, 20 MACH. INTELL. RES. 656 (2023).

<sup>118</sup> Jia-Chen Gu et al., *MPC-BERT: A Pre-Trained Language Model for Multi-Party Conversation Understanding*, in PROC. OF THE 59TH ANN. MTG. OF THE ASS'N FOR COMPUT. LINGUISTICS AND THE 11TH INT'L JOINT CONF. ON NATURAL LANG. PROCESSING 3682 (2021).

<sup>119</sup> David Bamman, Ted Underwood & Noah A. Smith, *A Bayesian Mixed Effects Model of Literary Character*, in PROC. OF THE 52ND ANN. MTG. OF THE ASS'N FOR COMPUT. LINGUISTICS 370 (2014).

<sup>120</sup> U.C. BERKELEY LIBRARY, AI & AI-friendly TDM Insert Language, available at <https://docs.google.com/document/d/12ENn4wDHXsuZpJyIYcqcrrGJIsIT28ImX0AYEmrhGU/edit?usp=sharing>; see also Rachael Samberg & Katie Zimmerman, *How Can We License AI Uses and Training Rights?*, in E-RESOURCE LICENSING EXPLAINED (2024), <https://berkeley.pressbooks.pub/eresourcelicensingexplained/chapter/artificial-intelligence-part-3-how-can-we-license-ai-uses-and-training-rights/>.

<sup>121</sup> COMMUNIA, *supra* note 14, at 18 (FinELib, which negotiates for libraries in Finland, could not reach agreement with American Chemical Society).

<sup>122</sup> McCracken & Raub, *supra* note 79, at 8 (“In addition to ProQuest and Gale, other vendors in this area include Elsevier, LexisNexis, JSTOR, and NewsBank. Given their challenges in negotiating copyright with a primary publisher, and their need to control the data mining process, vendors in this situation often provide the most challenging instances in which to work.”)

layer cake of license agreements stacking restrictive usage rights upon restrictive usage rights to the detriment of downstream libraries and scholars who may be precluded from securing the uses they need.

### C. The knowledge economy

One aspect of this landscape we have not yet addressed is the “why.” Why are publishers doing this? In some part, we believe publishers have attempted to curb TDM and AI usage because they are concerned about: (1) the security of their licensed products, and the fear that researchers will leak or release content behind their paywalls; and (2) AI being used to create a competing product that could substitute for the original licensed product and undermine their share of the market. While these concerns are valid, we believe they reflect longstanding fears over users’ potential generalized misuse of licensed materials. But publishers are already able to—and do—impose contractual provisions disallowing the creation of derivative products and systematically sharing licensed content with third-parties, so additionally banning the use of AI in doing so is unwarranted.<sup>123</sup> In all events, the TDM and AI-specific licensing language that libraries have created prohibits these outcomes and provides remedies to publishers in the event of breach.<sup>124</sup>

Another reason some publishers seek to restrict fair use TDM and AI in research is because publishers wish to impose monopoly-like control over the research lifecycle. These publishers are offering their own products and services—at an additional cost of several hundred thousand dollars a year, in some cases—to “lock in” researchers into the publishers’ own systems and platforms.<sup>125</sup> This would mean that, in addition to paying hundreds of thousands of dollars a year to license access to the content, libraries would

<sup>123</sup> See, e.g., License Agreement Between Springer-Nature and the Regents of the University of California, <https://ucop.app.box.com/s/v1260wmpe8pmm1jepwj36t7ulklcssi8> (Par. 3.1, “Neither Licensee nor Authorized Users shall... (ii) Except as permitted by applicable law or this License Agreement, update, change, revise, adapt, modify, translate, transform or create any derivative work of the Content;” TDM Product Terms Par. 4(ii), “Licensee and Authorized Users may not... create derivative products or services that would compete with Licensor or Licensor’s Affiliates products or services”); Planet Labs, Master Content License Agreement, [https://drive.google.com/file/d/19QaSCzKtw\\_v8nXY\\_HxJJ4uSx9uQyUEun/view?usp=sharing](https://drive.google.com/file/d/19QaSCzKtw_v8nXY_HxJJ4uSx9uQyUEun/view?usp=sharing) (prohibiting creation of derivative works).

<sup>124</sup> Rachael Samberg & Katie Zimmerman, *Artificial Intelligence Part 3: How Can We License AI Uses & Training Rights?*, in E-RESOURCE LICENSING EXPLAINED, <https://berkeley.pressbooks.pub/eresourcelicensingexplained/chapter/artificial-intelligence-part-3-how-can-we-license-ai-uses-and-training-rights/>.

<sup>125</sup> 11 J. LIBRARY. SCHOLARLY COMM’N., no. 1, at 8-9 (discussing fee-based vendor TDM portals); see also Authors Alliance, Text and Data Mining under U.S. Copyright Law: Landscape, Flaws & Recommendations, at 9-10 (Oct. 2024), <https://www.authorsalliance.org/wp-content/uploads/2024/11/Text-and-Data-Mining-Report-102024.pdf> (“On top of this, vendors have increasingly started using their own proprietary TDM platforms”); Sean M. Fiil-Flynn et al., *Legal Reform to Enhance Global Text and Data Mining Research*, 378 SCIENCE 951 (2022) (“Some publishers make limited copyright licenses available for TDM uses, often for additional fees charged to libraries or researchers. But paid licensing is not an affordable or viable option for many critical TDM projects.”)



also need to conjure hundreds of thousands of non-existent dollars each year to license access to the TDM and AI tools the publisher is selling to make fair uses of the underlying content.<sup>126</sup> This is precisely what the Constitution intended copyright law to discourage: the permanent entrenchment of unchecked control over the knowledge economy in ways that prohibit use and reuse.<sup>127</sup>

To that end, publishers' efforts to forbid TDM, or the training or dissemination of AI tools that now "know" something based on the licensed content, defy the purpose of licensing content. When research libraries license content for scholars, those scholars are learning information from the content.<sup>128</sup> When scholars study and write about the content, they are not regenerating the actual expression from the content, the part that is protected by copyright; rather, they are conveying the lessons learned from it—facts not protected by copyright.<sup>129</sup> One way to think about this problem is that prohibiting TDM and the training or dissemination of AI tools for TDM is functionally equivalent to prohibiting scholars from learning anything about the content libraries are licensing to them from which to learn. The basic tenets underpinning the 1976 Copyright Act uphold that copyright owners should not be able to monopolize the dissemination of information learned from their properly-acquired content, and especially when that information is used non-commercially for scientific research.<sup>130</sup> If the restrictive licensing trends continue unchecked, copyright owners will further erode the ability of researchers, educators, and students to freely engage with knowledge, curtailing freedom of inquiry and reinforcing economic inequality around scholarly communication.

<sup>126</sup> 11 J. LIBRARY. SCHOLARLY COMM'N., no. 1, at 4 (literature review of licensing practices and terms for text and data mining "notes Ann Okerson's 2013 message that 'librarians do not want to see a future where researchers (and libraries) must depend on costly publisher tools and services, in addition to the large sums we are already paying for e-resources' (p. 10). The future that understandably concerned Okerson and others has, we find, very much come to pass.")

<sup>127</sup> Tyler T. Ochoa & Mark Rose, *The Anti-Monopoly Origins of the Patent and Copyright Clause*, 84 J. COPYRIGHT SOC'Y U.S.A. 909 (2007); *Eldred v. Ashcroft*, 537 U.S. 186 (2003) (relying on *Bonito Boats*, 489 U.S. 141 (1989) (the grant of copyright is Constitutional because it reflects "a balance between the need to encourage innovation and the avoidance of monopolies which stifle competition without any concomitant advance in the 'Progress of Science and useful Arts'"); see *supra* note 17.

<sup>128</sup> See *Bartz v. Anthropic PBC*, No. C 24-05417 WHA, at 12 (N.D. Cal. June 23, 2025) ("Authors cannot rightly exclude anyone from using their works for training or learning as such. Everyone reads texts, too, then writes new texts. They may need to pay for getting their hands on a text in the first instance. But to make anyone pay specifically for the use of a book each time they read it, each time they recall it from memory, each time they later draw upon it when writing new things in new ways would be unthinkable. For centuries, we have read and re-read books. We have admired, memorized, and internalized their sweeping themes, their substantive points, and their stylistic solutions to recurring writing problems.")

<sup>129</sup> *Id.* See also 17 U.S.C. § 102(b) ("In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.")

<sup>130</sup> See *supra* note 17.

## II. CONSUMER MARKET MATERIALS

This section focuses on popular e-books and films that publishers or distributors license through paywalled-access platforms direct to consumers like researchers. Books and films of course are the types of content that libraries have historically collected and preserved, providing an important immediate resource for researchers today while also preserving materials so researchers long into the future can gain historical insights into culture at particular periods of time. For researchers, these popular materials—from NY Times best sellers to niche, cult-classic films—form an important resource for a wide variety of fields, from literature to history to information science. Researchers commonly rely on copyright limitations and exceptions to make use of these works—from simply showing a film, relying on Section 110(1) exception for classroom instruction,<sup>131</sup> to extensive reproduction for creating a corpus for computational analysis, relying on Section 107’s fair use provision.<sup>132</sup>

Unlike specialty research resources, for an increasingly large portion of popular cultural works, distributors license them direct to consumers through non-negotiable contracts known as End User License Agreements (EULAs). Libraries, for the most part, are unable to secure licenses under these standard terms, as we highlight through our examination of contract terms below, and distributors have for the most part refused to offer libraries any alternatives. For individual researchers and educators, personal subscriptions are common but other license terms in these EULAs cast significant doubt on their ability to engage in normal research and educational uses under the strict terms of these agreements.

In this section we first sketch out what the market for these materials looks like — which distributors control what kinds of works that researchers use, and how access to alternatives (for example, physical media) has evolved in recent years. This context is important to understand because it affects the ability of libraries and researchers to identify alternative, non-licensed access to these works (for example, by acquiring physical media instead, such as print books or DVDs). The concentration of distribution channels combined with increasing exclusivity of distribution largely means that libraries and research institutions are shut out from obtaining licensed access to these materials—i.e. the distributors will not even offer a license agreement to a library for its users. After that brief overview of the market, the remainder of this section reviews the clauses from EULAs of the largest e-book and streaming film distributors, explaining how five common types of clauses result in effectively limiting the types of uses libraries and the researchers and educators they serve can make of the materials subject to these EULAs.

---

<sup>131</sup> 17 U.S.C. § 110(1).

<sup>132</sup> See, e.g., D. Bamman et al., *Measuring diversity in Hollywood through the large-scale computational analysis of film*, PROC. NAT’L ACAD. SCI. U.S.A. 121 (46) e2409770121, <https://doi.org/10.1073/pnas.2409770121> (2024) (discussing reliance on Fair Use to conduct computational research on 2,307 films representing the top 50 movies by US Box Office performance from 1980 to 2022).

### A. The market for consumer works

When it comes to popular culture, the vast majority of e-books or streaming movies sold in the United States are governed by one of twelve non-negotiable EULAs.<sup>133</sup> Five companies (and five contracts) control approximately 95% of all US e-book purchases in the United States,<sup>134</sup> and another seven (and their seven contracts) control 98% of all streaming media purchases in the United States.<sup>135</sup> These non-negotiable licensing terms are drafted by dominant market players (with legal teams of hundreds of lawyers) such as Amazon, Apple, or Netflix.

It is worth observing the scale at which this small handful of publishers and distributors dominate the market for popular creative works. Obtaining accurate market share statistics for e-book and audiobook vendors is challenging; private companies aren't required to disclose this information. Their selective release of sales figures provides strategic advantages for publishers, and market shares can also change quickly. For example, Rakuten Kobo's share grew when it partnered with Walmart, while Barnes & Noble declined in the late 2010s due to closing stores and phasing out its Nook e-reader.<sup>136</sup>

But a few facts about digital book sales can be readily discerned. First, Amazon dominates the trade e-book market, and has since that market's inception.<sup>137</sup> Second, other major players—which collectively have a minority share of the e-book market—include Barnes & Noble, Apple, Google, Rakuten Kobo, ebooks.com, and Smashwords.<sup>138</sup> Other market participants include some of the same companies dominating the e-book market, like Google and Kobo, as well as streaming services like Spotify and more specialized companies like Libro.fm, which partners with brick and

<sup>133</sup> We use the term “sold” loosely. The works offered for “sale” are licensed for access and use with significant strings attached. See Cal. A.B. 2426, 2023-24 Reg. Sess. (2024) (California law governing use of the term “sold” in this context); see also Aaron Perzanowski & Jason Schultz, *The End of Ownership* (2018).

<sup>134</sup> See Max Lakin, *Which Ebook Publishing Platform Is Best?*, MAGNOLIA MEDIA NETWORK/Magnolia Media Network, (March 20, 2020), <https://magnoliamedianetwork.com/ebook-publishing-platforms/>.

<sup>135</sup> Nielson, *The Gauge* (Apr. 2025), <https://www.nielsen.com/data-center/the-gauge/#viewing-by-distributor> (showing streaming views occupying 44.3 of all TV viewership and seven firms accounting for over 98% of those streams).

<sup>136</sup> Andrea Cheng, *Why Walmart Is Pushing Into E-Books, A Business On The Decline*, Forbes, Aug. 22, 2018, <https://www.forbes.com/sites/andriacheng/2018/08/22/walmart-introduces-ebooks-in-its-latest-lou-d-display-of-intense-fight-against-amazon/>.

<sup>137</sup> For a discussion on how the monopoly power of Amazon could make even copyright law irrelevant, see Guy Rub, *Amazon and the New World of Publishing*, 14 ISJLP 367 (2018).

<sup>138</sup> See *E-Book Market – Growth, Trends, Covid-19, Impact and Forecasts (2023-2028)*, MORDOR INTEL., <https://www.mordorintelligence.com/industry-reports/e-book-market>; *Where to Buy EBooks, MAKE USE OF* (Jul. 14, 2022), <https://www.makeuseof.com/tag/buy-ebooks-best-ebook-stores/>. Things are just as unclear for audiobooks, but after buying Audible.com in 2008, we observe that Amazon largely took over this market too.

mortar bookstores to allow readers to purchase digital audiobooks from those bookstores. In the realm of streaming services, the major market participants are identifiable based on advertising and the popularity of works available on their platforms. These major players include Netflix, Hulu, HBO, YouTube, Disney, Amazon, and Peacock.<sup>139</sup> For both types of works, it is clear that a very small number of contracts govern the overwhelming majority of popular digital books and films.

As discussed above, one historical pathway for libraries and researchers to avoid contractual limits on their use has been to secure physical copies of works for circulation and archiving, or digitize or reproduce them for reuse. For some consumer-market materials, such as e-books, it is often the case that a physical copy is still available though in certain contexts that norm is being challenged. But digitization is often not a viable option for other reasons, primarily cost and quality of the resulting scans. The situation is especially dire for audiobooks, where Amazon exerts exclusive control over the majority of titles; physical CDs are rarely available for purchase.<sup>140</sup>

For film (and other works such as music, not discussed in detail here), it has become increasingly difficult to source non-streaming copies. Physical distribution has seen a dramatic and increasing decline in availability.<sup>141</sup> Some streaming services that specialize in the academic market, such as Kanopy and Swank, cater to academic libraries and educational institutions, but their catalogs offer only a subset of the broader materials that one might find available through the most popular streaming platforms. This is especially true because increasingly, these platforms host their own exclusive content. Netflix, just to pick one, exemplifies this: as of this writing, the Netflix catalog for its U.S. users included some 7793 titles, 4,427 (over 55%) of which are Netflix originals that are for the most part exclusively available through the Netflix streaming platform.<sup>142</sup>

<sup>139</sup> See *Top Streaming Services by Subscribers*, FLIXPATROL, <https://flixpatrol.com/streaming-services/subscribers/>.

<sup>140</sup> Caitlin Marsh, *The Problem with Exclusives*, BRILLIANT BOOKS (Apr. 8, 2021), <https://www.brilliant-books.net/problem-exclusives>.

<sup>141</sup> See DIGIT. ENT. GRP., *U.S. Consumer Spending on Digital Entertainment Consumed at Home and on the Go Grew \$10 Billion in 2024, Surpassing \$57 Billion*, DIGIT. ENT. GRP. (Feb. 14, 2025), [https://www.degonline.org/wp-content/uploads/2025/02/fl\\_Year-end-2024-Digital-Media-Entertainment-report-cover-note-with-grid.pdf](https://www.degonline.org/wp-content/uploads/2025/02/fl_Year-end-2024-Digital-Media-Entertainment-report-cover-note-with-grid.pdf) (reporting a dramatic 24% decline in year over year physical media sales, which already constituted a small portion of overall digital entertainment sales as compared to streaming and digital distribution).

<sup>142</sup> WHAT'S ON NETFLIX, <https://www.whats-on-netflix.com/library/> (last updated Jan 1, 2025). Netflix does make a handful of its Netflix Originals titles available on DVD. See Kasey Moore, *List of Netflix Originals Available on DVD & Blu-ray*, WHAT'S ON NETFLIX (Feb. 16, 2021), <https://www.whats-on-netflix.com/news/list-of-netflix-originals-available-on-dvds-blu-rays/>. Additionally, this analysis can be somewhat complicated because some Netflix Originals are only on Netflix for a short time before being sold off or relicensed via another distributor. So we acknowledge some titles may become available in physical format in other ways. See Kasey Moore, *Every Netflix Original Movie and Series That's Been Removed From The Streamer (And Why)*, WHAT'S ON NETFLIX (Apr. 8, 2025), <https://www.whats-on-netflix.com/leaving-soon/every-netflix-original-removed-from-netflix/>.

These factors are important for libraries and researchers when it comes to the issue of contractual override for two reasons: (1) they mean that even small changes in EULAs can impact a huge number of copies of works, and (2) the scale of these vendors coupled with an increasing move toward not just distributing content but securing exclusive distribution rights means that many materials are *only* available through these vendors under the terms that they dictate.

*B. How consumer contracts limit lawful use for libraries*

These consumer market EULAs and the environment in which they arise are quite different from those directed at academic institutions and libraries in a few key ways. First, EULAs are contracts of adhesion: non-negotiable and offered to the public as a take-it-or-leave-it proposition. The terms are not only dictated by the distributor, they are also typically subject to change based on the unilateral decision of the seller, making it difficult to establish with certainty the future legality of uses obtained from a vendor.<sup>143</sup> Libraries have essentially no opportunity to secure stable licensing terms that reflect their needs: through our own experiences as well as in working with librarians and others in drafting this paper, we know that attempts to secure modifications from Amazon, Netflix, Apple, and others to allow for preservation, TDM use, public performance rights and others are not successful. When librarians have received a response, it was that such changes to EULAs were not possible.

Second, we see several themes across these licenses in the ways that they restrict downstream uses otherwise permitted by copyright law, for both consumers as well as libraries and scholars. For this article, we examined the EULAs of some of the top e-book distributors and streaming film services.<sup>144</sup> These include:

- Amazon’s “Kindle Store Terms of Use,” [“Amazon Kindle”]
- Apple’s “Apple Media Services Terms and Conditions,” [“Apple”]
- Google Play’s “Google Play Terms of Service,” [“Google Play”]

<sup>143</sup> A. Chakraborty, R. Shankar & J.R. Marsden, *An Empirical Analysis of Consumer-Unfriendly Ecommerce Terms of Service Agreements: Implications for Customer Satisfaction and Business Survival*, 53 ELEC. COM. RES. & APPLICATIONS, Mar.-Apr. 2022, at 101151 (“a key element in most ToS is the claim that the e-retailer can change the ToS at any time (even after a sale) and that the customer is bound by any and all altered conditions. Given that customers do not appear willing to actually take the time to carefully read ToS, it is very unlikely that the customer would continually read and check a ToS for updates.”) See, e.g., Apple Media Services Terms and Conditions, APPLE, <https://www.apple.com/legal/internet-services/itunes/> (last updated Sept 16, 2024) (“Apple reserves the right at any time to modify this Agreement and to add new or additional terms or conditions on your use of the Services. Such modifications and additional terms and conditions will be effective immediately and incorporated into this Agreement. Your continued use of the Services will be deemed acceptance thereof.”).

<sup>144</sup> We would like to thank UC Berkeley law students Billy Malmed, Lawrence Myung, Sophie Zollman, and their clinical supervisor in the Samuelson Law, Technology and Public Policy Clinic, Brianna Schofield, for background research support and their excellent background memo on which we draw for this analysis. We also thank for Authors Alliance Senior Staff Attorney Rachel Brooke for her contributions.

- Barnes & Noble (which has two licenses governing e-book content—the “Nook Device Terms of Service” [“Nook TOS”] and the “Digital Content Terms of Sale” [Barnes & Noble Digital TOS])
- Amazon’s “Amazon Prime Video Terms of Use,” [“Amazon Prime”]
- Disney+’s “Disney+ and ESPN+ Subscriber Agreement,” [“Disney”]
- HBO’s “HBO Max Terms of Use,” [“HBO”]
- Netflix’s “Netflix Terms of Use,” [“Netflix”]
- YouTube’s “YouTube Paid Service Terms of Service” and “Terms of Service” [“YouTube”]<sup>145</sup>

What follows is a brief review of the many different ways that these contracts act to limit library, researcher, and educational user rights under fair use and other limitations and exceptions.

1. Redefining the universe of acceptable use: blanket prohibitions on reproduction, adaptation, and distribution

One of the most important ways that EULAs limit users’ ability to engage in uses covered by copyright limitations and exceptions is to create contractual obligations for users that limit their ability to reproduce, adapt and redistribute materials in *any* way beyond that which is necessary to view or make carefully prescribed uses of the the licensed content. In doing so, these EULAs essentially extract a promise from the user to not exercise the flexibility that copyright limitations and exceptions otherwise afford them.

Of course, copyright owners hold rights under Section 106 to control reproduction, distribution, derivative works, and public performance and display. It may seem, then, that copyright owners are merely restating their Section 106 rights in these EULAs.

<sup>145</sup> For simplicity we cite each of these EULAs here and refer back to them in the text by their distributor name, indicated in brackets above. See *Kindle Store Terms of Use*, AMAZON [“Amazon Kindle”] (last updated Dec. 31, 2024), <https://www.amazon.com/gp/help/customer/display.html?nodeId=201014950>; *Apple Media Services Terms and Conditions*, APPLE LEGAL [“Apple”] (last updated Sep. 16, 2024), <https://www.apple.com/legal/internet-services/itunes/>; *Google Play Terms of Service*, GOOGLE [“Google Play”] (last updated Mar. 15, 2023), [https://play.google.com/intl/en-US\\_us/about/play-terms/index.html](https://play.google.com/intl/en-US_us/about/play-terms/index.html); *NOOK Device Terms of Service*, BARNES & NOBLE [“Nook TOS”] (last updated May 4, 2018), <https://www.barnesandnoble.com/h/nook-device-terms-of-service>; *Digital Content Terms of Sale*, BARNES & NOBLE [Barnes & Noble Digital TOS] (last updated Sep. 16, 2024), <https://www.barnesandnoble.com/h/digital-content-terms-of-sale>; *Amazon Prime Video Terms of Use*, AMAZON [“Amazon Prime”] (last updated May 15, 2025), <https://www.primevideo.com/help?nodeId=202095490&view-type=content-only>; *Disney+ and ESPN+ Subscriber Agreement*, DISNEY+ [“Disney”] (last updated May 5, 2025), <https://www.disneyplus.com/legal/subscriber-agreement>; *HBO Max Terms of Use*, HBO [“HBO”] (last updated Oct. 28, 2024), <https://www.max.com/terms-of-use/en-us>; *Netflix Terms of Use*, NETFLIX [“Netflix”] (last updated Apr. 17, 2025), <https://help.netflix.com/legal/termsofuse>; *YouTube Paid Service Terms of Service*, YOUTUBE (last updated Apr. 14, 2025), [https://www.youtube.com/t/terms\\_paidservice](https://www.youtube.com/t/terms_paidservice); *Terms of Service*, YOUTUBE [“YouTube”] (last updated Dec. 15, 2023), <https://www.youtube.com/static?template=terms>.

Section 106, however, is carefully limited by its introductory clause that makes clear those rights subject to all of the limitations contained in Sections 107-122 that follow.<sup>146</sup> It might be sensible, then, for distributors in their EULAs to ask users to promise that they will not make uses that infringe copyright, or even a more specific promise to not violate the copyright owners rights under Section 106 of the Copyright Act. Either approach would ensure that the copyright owners rights are protected while still incorporating the whole host of limitations and exceptions that are inherent to the scope of Section 106 rights.

What we see instead in many consumer market EULAs is a subtle but critically different approach. In many EULAs we find terms that essentially redefine the universe of acceptable use, extracting promises from users to not engage in a wide variety of specific uses such as “reproduction” or “adaptation,” but then strictly limiting exceptions to that general rule. In doing so, these EULAs—whether intentionally or not—place a large number of classic and lawful reuses in a zone of uncertainty. On their face, these licenses transform routine fair uses such as making short quotations for commentary in a scholarly work or combining excerpts of works together for comparison or analysis into activities that are technically a breach of contract. For research library applications such as preservation, these limits are particularly problematic when one considers all the copying, decryption, file transformations, and other processing necessary to maintain the viability of digital content for the future.

The ways these EULAs do this vary only slightly between the versions. For example, for e-books, Apple’s license states that, “No portion of the Content or Services may be transferred or reproduced in *any* form or by *any* means, except as expressly permitted by this Agreement. You agree not to modify, rent, loan, sell, share, or distribute the Services or Content in *any* manner.” (emphasis added). Google Play’s license restricts users from selling, redistributing, modifying, or transferring “*any* Content to *any* third party . . . except as specifically permitted and only in the exact manner provided.” (emphasis added). Similarly, for streaming media, Netflix tells users that “[e]xcept as *explicitly authorized by us*, you agree not to: (i) archive, reproduce, distribute, modify, display, perform, publish, license, create derivative works from, offer for sale, or use content and information contained on or obtained from or through the Netflix service.” (emphasis added)

One of the key elements in all of these clauses is that the EULA asks the user to promise not to engage in *any* of a wide variety of uses *except* for a small list of permitted activities. In none of these agreements is there any recognition that limitations and exceptions may apply to certain activities, or that the blanket prohibitions (e.g., prohibiting reproduction) may in fact have exceptions not listed explicitly in the contract. While some of these terms are easier to see their direct conflict with library and research uses, such as Netflix’s explicit ban on making archival copies (allowed by Section 108 as well as Section 107), they all effectively create the same problem: library and research users are induced to make a promise, by entering into this EULA, to not make a wide variety of uses that the Copyright Act says they ought to be able to make

<sup>146</sup> 17 U.S.C. § 106 (2025) (“Subject to sections 107 through 122, the owner of copyright under this title has the exclusive rights to do and to authorize [the rightsholders enumerated exclusive rights.]”).



unless the licensor explicitly permits it. These clauses effectively redefine the universe of acceptable use into one dictated by the licensor rather than the Copyright Act.

There are a few exceptions to this general approach. Barnes & Noble, for instance, has two licenses governing e-book content—the “Nook Device Terms of Service” and the “Digital Content Terms of Sale.” Because these licenses have conflicting terms, it is ambiguous whether Barnes & Noble puts boundaries on a licensee’s uses in a way that restricts users’ ability to make fair uses. The Nook TOS grants a license only to “access and make personal, non-commercial use” of e-books, and also has an extensive list of prohibited conduct, such as “you may not transfer, copy, manipulate, or display” e-books. The Nook TOS also states that users may not “upload, post, reproduce, or distribute in *any* way” e-books without permission of the copyright owner. However, the Nook TOS explicitly includes a fair use provision: “Except as permitted by the fair use privilege under the U.S. copyright laws (see, e.g., 17 U.S.C. Section 107) or similar provisions in applicable local intellectual or copyright laws and treaties.” The Barnes and Noble Digital TOS also states that users are restricted to “personal, non-commercial use” (a common limitation discussed in the next subsection) and cannot “upload, post, reproduce, or distribute in *any* way” e-books, without the copyright owner’s permission. (emphasis added). However, unlike the Nook TOS, the Digital TOS does not include a provision that allows for uses that would be permitted under fair use. The fact that it is not included doesn’t necessarily mean there is contract override, but does leave the question lingering. The key differences between these two licenses—the Nook TOS’s acknowledgement of licensee’s fair use rights and the Barnes & Noble Digital TOS’s silence on this point—creates ambiguity.

Google Play offers another variation, providing licensees some leeway when it comes to the public display of licensed content. Although the license states a licensee may not publicly display any licensed content, it provides for an exception “where such use would not constitute a copyright infringement or violate any other applicable right.” By offering an allowance for displaying content in a non-infringing manner, researchers or libraries may be able to make uses of licensed content that implicate the display right. That said, this provision is limited in two ways. First, it is not clear whether a licensee is allowed to display content for personal and non-commercial purposes *only*, or if *any* display of content not in violation of copyright law is acceptable, because elsewhere the license restricts user’s display rights to “personal, noncommercial use[s].” In traditional academic settings, such as in the classroom or as part of an academic symposium, this conflict may seem unlikely to arise. But many scholars seek wider audiences for their work including paid and compensated presentations at corporate events and conventions. A researcher who seeks to display their research in those settings, even when just using a very short excerpt that they use appropriately to comment or criticize, may find themselves technically in breach of their Google Play EULA. Second, public display is only one of the many types of uses an academic author or research library would likely seek to make in their work, with many others implicating reproduction and distribution rights.

EULAs that impose blanket prohibitions (“you may not make *any* reproduction”) with only narrow, contractually-designated, explicit exceptions, can lead to absurd results if taken at face value. Indeed, it seems unlikely that most distributors would attempt to enforce them strictly for many of the kinds of everyday uses one might be

surprised to find a conflict with an academic quoting extensively from a e-book to write a book review, for instance (a reproduction that is not explicitly allowed under the terms, and so technically a breach of the EULA). We admit that for the most part, these uses are not of great concern for practical reasons. But larger, more systematic uses that depend on long-term stability—a large-scale text data mining research project, or large-scale library preservation, for instance—pose a more significant problem. In those cases, libraries and researchers are understandably hesitant to invest the significant time, infrastructure, and cost necessary when the licensor could at any moment choose to strictly enforce contractual rights that would undermine such projects. Copyright law and its limitations and exceptions give those users certainty about the future of their uses while existing EULAs undermine that certainty.

## 2. Restricting any use that is not personal and non-commercial

A second major way that EULAs limit user behavior covered by copyright exceptions is by asking users to agree that their uses will *only* be for personal, non-commercial uses. At first glance, EULAs that restrict commercial use may seem just fine. The fair use analysis, after all, counts commercial use as a negative,<sup>147</sup> and the Supreme Court in *Andy Warhol Foundation v. Goldsmith* recently reinforced that commercialism is an important part of both the first and fourth factor fair use analysis, stating that a transformative use “must be balanced against the commercial nature of the use.”<sup>148</sup>

Lower courts, however (including many courts post-*Warhol*), have continued to find that while commercialism is an important factor, it is not outcome determinative. Fair use still applies in the commercial context. Recent cases have approvingly found that a wide variety of commercial uses are fair, such as for-profit educational kits that incorporated an artist’s copyrighted image,<sup>149</sup> commercial use of copyrighted home floor

---

<sup>147</sup> 17 U.S.C. § 107(1) (2025).

<sup>148</sup> *Andy Warhol Foundation Visual Arts v. Goldsmith*, 598 U.S. 508 (2023). The court goes on to explain the particular context in which the commercial inquiry is most important: “[i]f an original work and a secondary use *share the same or highly similar purposes*, and the secondary use is of a *commercial nature*, the first factor is likely to weigh against fair use, absent some other justification for copying.”

<sup>149</sup> *Keck v. Mix Creative Learning Center, LLC*, 116 F. 4th 448, 454 (5th Cir. 2024) (“Although Mix Creative is a commercial enterprise, its use of Keck’s copyrighted works was transformative, as the art kits had an educational purpose that was significantly different from the original, decorative purpose of Keck’s dog-themed artworks.”)

plans in real estate listings.<sup>150</sup> In fact, fair use approves of many commercial uses, and even the copying of millions of works for AI training.<sup>151</sup>

Many well-accepted fair uses, even for research and library uses, are not personal or non-commercial, and thus fall outside of these categories.<sup>152</sup> Libraries, as institutional and not personal users, are almost entirely excluded from making any uses under these licenses, including ones permitted by fair use or other limitations and exceptions. This would include ordinary exercises of fair use, for example, in making excerpts of chapters to place on course e-reserves in support of teaching.<sup>153</sup> The “commercial” restriction also means that all sorts of routine fair uses—e.g., copying a portion of a video to create a parody that might be sold on a commercial basis—would be considered to be a breach of the EULA.<sup>154</sup> Virtually every contract noted above has such a limitation, with almost identical language, and with no recognition of any exceptions. For the Amazon Kindle, for instance, users must agree that they will use such content “*solely* for [their] your personal, non-commercial use.” Apple says users may only use e-books “*only* for personal, noncommercial purpose” and cannot be used “in any way whatsoever except . . . for personal, noncommercial uses.” Google says that e-books are for “personal, non-commercial use *only*.” Netflix states that “The Netflix service and any content accessed through our service are for your personal and non-commercial use only and may not be shared with individuals beyond your household unless otherwise allowed by your subscription plan” (Netflix offers no institutional plan for libraries or academic institutions). HBO provides that “[y]ou are *only* authorized to access and display Content on the Platform in the Service Area for your non-commercial, personal, entertainment use.”

These clauses have for the most part meant that libraries cannot make any meaningful use of these licensed resources because they are institutional users, and not personal users. On occasion some have attempted, for example, by purchasing Netflix

<sup>150</sup> *Designworks Homes, Inc. v. Columbia House of Brokers Realty, Inc.* 126 F.4th 589, 597 (8th Cir. 2025) (“That the agents incorporated the floorplans in commercial advertising counterbalances the transformativeness of their use of the designs only in part. Commercial uses of a work are less favored than noncommercial uses, and we view commercial advertising uses with some skepticism. . . . But the commerciality of a use is troubling primarily when the use displaces the work or derivative works in the market.”).

<sup>151</sup> *Bartz v. Anthropic PBC*, Case No. No. C 24-05417 WHA, Order on Fair Use (N.D. Cal. June 23, 2025) (“Yes, Anthropic is a commercial outfit. And, this order takes for granted that Anthropic in fact benefited from the print-to-digital format change — or it would not have gone to all the trouble. But the crux of the first fair use factor's concern for “commercial” use is in protecting the copyright owners and their entitlements to exploit their copyright as they see fit (or not). . . . That the accused is a commercial entity is indicative, not dispositive. That the accused stands to benefit is likewise indicative. But what matters most is whether the format change exploits anything the Copyright Act reserves to the copyright owner.”).

<sup>152</sup> See, e.g., *New Era Publ’ns Int’l Aps v. Carol Publ’g Grp.*, 904 F.2d 152 (2d Cir. 1990); *Wright v. Warner Books, Inc.*, 953 F.2d 731 (2d Cir. 1991).

<sup>153</sup> See *Cambridge Univ. Press v. Albert*, 906 F. 3d 1290 (11th Cir. 2018) (university library uses for e-reserves was fair use in many but not all cases, particularly where amounts used are limited and electronic licensed copies are not available for the library to purchase).

<sup>154</sup> *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994) (commercial parody held to be fair use).

licenses to provide access for their uses,<sup>155</sup> making the argument that access is still just for one individual at a time and therefore “personal.” But the prevailing view, reinforced by Netflix,<sup>156</sup> is that such uses are prohibited by the contract. In that particular scenario, the type of access that libraries hoped to provide—full streaming of film to users—would have required licensed permission anyway; fair use would likely not apply to allow libraries to stream full films to their users.<sup>157</sup> But, the “personal use only” restriction sweeps away a wide variety of other uses, such as making institutional preservation copies that the Copyright Act allows under Section 108(b) and (c), excerpts for course reserve use that fair use allows for under Section 107. Libraries who attempt to make such uses may not face a copyright infringement claim, but are likely to find themselves in breach of the EULA at issue.

For researchers, these EULA restrictions can also impose barriers. For example, imagine an academic author who quotes from a licensed e-book in a scholarly book that the author then sells commercially and receives royalties. Or a historian who includes excerpts of a film for use in an educational documentary that is meant to be sold commercially. None of the EULAs we examined contained a specific definition of “non-commercial.”<sup>158</sup> At least in the context of fair use, some courts have concluded that

<sup>155</sup> Adrienne Jeffries, *Netflix Turns a Blind Eye to Illegal Use by School Libraries*, READWRITE ReadWrite, (Sept. 20, 2010),

[https://readwrite.com/netflix\\_turns\\_a\\_blind\\_eye\\_to\\_illegal\\_use\\_by\\_school/](https://readwrite.com/netflix_turns_a_blind_eye_to_illegal_use_by_school/).

<sup>156</sup> Travis Kaya, *Academic Libraries Add Netflix Subscriptions*, Chronicle of Higher Education, September 18, 2010,

<https://www.chronicle.com/blogs/wiredcampus/academic-libraries-add-netflix-subscriptions>

(“[A]s more librarians sign up for its popular mail and streaming-video services, Netflix says library distribution of rented DVD’s or streaming video violates its terms of use.”)

<sup>157</sup> There are few cases on point on this in the educational streaming context, but at least one court, in *Association for Information Media and Equipment v. The Board and Regents of the University of California*, Case No. 2:10-cv-09378-CBM (MANx) (C.D. Cal. 2012), concluded UCLA’s efforts to stream entire films to students enrolled in classes where those films were assigned and where UCLA had lawful access to the films, created “at a minimum, ambiguity as to whether Defendants’ streaming constitutes fair use and that it would not have been clear to a reasonable person in Defendants’ position that its streaming did not constitute fair use.”)

<sup>158</sup> See Creative Commons, *Defining “Noncommercial”*: A Study of How the Online Population Understands “Noncommercial Use” (2010),

[https://mirrors.creativecommons.org/defining-noncommercial/Defining\\_Noncommercial\\_fullreport.pdf](https://mirrors.creativecommons.org/defining-noncommercial/Defining_Noncommercial_fullreport.pdf)

(reporting that across several thousand focus group participants and survey respondents, “The empirical findings suggest that creators and users approach the question of noncommercial use similarly and that overall, online U.S. creators and users are more alike than different in their understanding of noncommercial use. Both creators and users generally consider uses that earn users money or involve online advertising to be commercial, while uses by organizations, by individuals, or for charitable purposes are less commercial but not decidedly noncommercial.”)

uses such as these may well be commercial, but are still protected by fair use.<sup>159</sup> In our experience in working both with a number of university general counsel, librarians, and counsel on the other side for these vendors, most understand these restrictions on commercial use to disallow such uses because they are in violation of the EULA.<sup>160</sup>

### 3. Prohibitions on public performance

Academic and library users have special permissions under U.S. copyright law to make certain public displays and performances of copyrighted works without permission under Sections 107, 108, and 110. The Copyright Act's public performance and display rights are broad. Performance, for instance, encompasses a wide range of activities: "To **perform**" a work means to recite, render, play, dance, or act it, either directly or by means of any **device** or **process** or, in the case of a motion picture or other audiovisual work, to show its images in any sequence or to make the sounds accompanying it audible."<sup>161</sup> Likewise, "public" performance or display is broad: "at a place open to the public or at any place where a substantial number of persons outside of a normal circle of a family and its social acquaintances is gathered."<sup>162</sup>

It seems clear that Congress recognized that many long-accepted classroom, educational, and non-profit public performances uses would be swept in under these broad sets of rights,<sup>163</sup> and so in addition to the flexibility limitation provided by Section 107, Congress enacted a number of specific exceptions contained within Section 108 and 110.

<sup>159</sup> See e.g., *Sundeman v. Seajay Society*, 142 F.3d 194, 203 (4th Cir. 1998) (in the context of a fair use analysis, concluding that a historical society that copied a manuscript to facilitate scholarship was engaged in non-commercial use but that the scholar who made and used copies to write scholarly analysis was engaged in commercial use because "there was a potential commercial motivation in that Dr. Blythe may have received royalties if her paper were published.")

<sup>160</sup> The only exception we could find among those EULAs we examined was Amazon Prime, which includes similar but less restrictive language: Amazon Prime Video's license affirmatively grants users a nonexclusive right to use licensed content for personal, non-commercial uses, but does not expressly limit a licensee to "only" these uses.

<sup>161</sup> 17 U.S.C. § 101.

<sup>162</sup> *Id.*

<sup>163</sup> For a helpful review of the public performance right in the educational context from before the advent of mass licensed works, see Francis M. Nevins, *Copyright, Cassettes and Classrooms: The Performance Puzzle*, 43 J. Copyright Soc'y U.S.A. 1 (1995). It's interesting to note Nevins's contemplated scenarios where copyright owners may seek to override Section 110(1), stating that, "It seems equally clear that if a studio tried to do an end run around the exemption, for example by printing on cassettes or their boxes a notice that purchasers and renters may not use the film in a classroom, a court would consider the notice a contract of adhesion that, being in conflict with a user's right under 110(1), could not be enforced." Nevins argued that such limits would not be enforced due to copyright preemption, based on the then-leading case on the issue, *Vault Corp. v. Quaid Software Ltd.*, 847 F.2d 255 (5th Cir. 1988). As we explain in Section IV, preemption may well be a viable argument against enforcement, though there are several countervailing precedents that make it far from certain.

Section 110(1) of the Copyright Act states that it is not an infringement to “perform[e] or display . . . a work by instructors or pupils in the course of face-to-face teaching activities of a nonprofit educational institution, in a classroom or similar place devoted to instruction,” but almost all of the streaming service EULAs would prohibit such uses.<sup>164</sup> Section 110(2) accommodates online teaching by allowing for “the performance of a nondramatic literary or musical work or reasonable and limited portions of any other work, or **display** of a work in an amount comparable to that which is typically displayed in the course of a live classroom session” subject to numerous institutional and instructional requirements.<sup>165</sup> Section 108, specific to library uses, allows libraries rights to display and perform a wide variety of works, subject to some limitations, when they are in their last 20 years of copyright protection. These exceptions are relied upon everyday by educators and librarians to engage in basic teaching and education – showing a film in class, exhibiting reproductions of archival materials physically and digitally, hosting a screening on campus for discussion. These are all uses that would otherwise be prohibited public performances or displays but for the 110 and 108 exceptions.

Almost all the major streaming platforms have terms that restrict public performance, most prohibiting it in *any* way.<sup>166</sup> For example, Netflix states that “You agree not to use the service for public performances.” Disney’s agreement provides that “You agree that you will not nor permit another person to do any of the following without our express written permission, and that these restrictions are a condition to your license: rebroadcast, transmit or perform the Content available via the Services.” HBO’s EULA requires users to promise that they will make *only* for the subscribers’ “non-commercial, personal, entertainment use,” and that the “**you may not** . . . display [or] perform publicly” any of its content (emphasis in original). The result of these flat prohibitions on public performance, with no acknowledgement of exceptions such as those contained in Section 110, means that a wide variety of educational and research uses that are permitted by the Copyright Act would represent a breach of the EULA.

One might think these flat prohibitions on public performance may be unintentional in sweeping so far. But experience and actions of these vendors indicate otherwise. Netflix, for instance, demonstrates this by offering a limited grant of permission for educational screenings for a small number of Original titles (a few hundred titles), which it describes as available for “one time” screenings, which Netflix defines to mean “that you can’t hold screenings several times in one day or one week - but if, for example, you are an educator who wants to show these films or series once a semester over multiple semesters, that’s okay.”<sup>167</sup> Netflix’s other main limitation on such educational use is that

---

<sup>164</sup> 17 U.S.C. § 110(1)-(2).

<sup>165</sup> 17 U.S.C. § 110(2).

<sup>166</sup> The YouTube Paid Service license, which provides a limited exception for uses involving public presentations. While the license prohibits using YouTube Premium content, in whole or part, in connection with a public presentation, it provides for an exception, “where such use would not constitute a copyright infringement.”

<sup>167</sup> Educational Screenings of Documentaries, Netflix, <https://help.netflix.com/en/node/57695> (last accessed June 11, 2025).

they must be “non-profit and non-commercial.”<sup>168</sup> The existence of this special license makes clear that Netflix understands that its main subscription offering does not extend to classroom or educational uses, and in preparation for the workshop that precipitated this paper, feedback from librarians indicated in every outreach to Netflix about classroom use has been consistent: such uses are not permitted under its EULA. For many users, this kind of restriction is surprising, and so libraries, which are often the primary source of information on campus about copyright and use of materials in classroom settings, have found themselves in the unenviable role of having to tell instructors that they cannot, because of these streaming services’ EULAs, do what the law otherwise allows them to do.<sup>169</sup>

#### 4. Restrictions on circumventing TPMs

Section 1201 of the Copyright Act provides that “[n]o person shall circumvent a technological measure that effectively controls access to a work protected under this title.”<sup>170</sup> “Circumvention of a technological measure” is in turn defined to mean “to descramble a scrambled work, to decrypt an encrypted work, or otherwise to avoid, bypass, remove, deactivate, or impair a technological measure, without the authority of the copyright owner.”<sup>171</sup> These provisions were enacted to stem digital piracy by attaching legal liability to efforts to reproduce or distribute digital works without the authorization of the rightsholder.<sup>172</sup>

<sup>168</sup> *Id.* This means that for-profit public performances would also not be permitted, which is consistent with Section 110(1) which only applies to nonprofit educational institutions.

<sup>169</sup> See, e.g., Tulane University Libraries, Film and Audio Resources @ Tulane University Libraries, <https://libguides.tulane.edu/c.php?g=1335135&p=9834699> (last updated April 22, 2025) (“Your personal Netflix account does not permit sharing or campus screenings. There are no options for academic libraries to purchase a campus screening license, either.”); Yale University Library, “Can I show Netflix movies in the classroom without seeking permission from the copyright owner?”, <https://ask.library.yale.edu/faq/175270> (last updated May 31, 2023) (“In general, **NO**. Movies fall into a different category than written works. Unless the movie is in the public domain, generally a license to view the movie will restrict public performances. A Netflix subscriber will have to abide by their Terms of Use (<https://www.netflix.com/TermsOfUse>), which explicitly say that viewing is for “personal” use only and “You agree not to use the service for public performances.” So, even though there might be a classroom exception in copyright law, the license (and your agreement to it) will prevail over the law.”); The University of Michigan-Flint, Media in the Classroom, <https://libguides.umflint.edu/faculty/media> (last updated June 1, 2025) (“Most commercial streaming services have clauses in their Terms of Use, Subscriber Agreement, or Terms of Service that prohibit the broadcasting of their content in any other setting than private and personal. Meaning that streaming content from these services either in a physical classroom or in an online classroom violates these agreements. The following list is a sample of the clauses from several popular streaming platforms, if you are considering using content from a service that isn’t listed, please consult their license or terms of use before showing any content in the classroom.”).

<sup>170</sup> 17 U.S.C. § 1201(a)(1)(A) (2025).

<sup>171</sup> 17 U.S.C. § 1201(a)(3) (2025).

<sup>172</sup> The Digital Millenium Copyright Act, S. Rpt. 105-190 at 8 (1998) (“Due to the ease with which digital works can be copied and distributed worldwide virtually instantaneously, copyright



Recognizing that Section 1201 could impose burdens on uses where circumvention may be necessary to make lawful uses, Congress at the same time sought to provide clarity for end users in cases where their uses were otherwise lawful.<sup>173</sup> Section 1201(a)(1)(B) provides that its prohibition on circumvention “shall not apply to persons who are users of a copyrighted work which is in a particular class of works, if such persons are, or are likely to be in the succeeding 3-year period, adversely affected by virtue of such prohibition in their ability to make noninfringing uses of that particular class of works under this title.”<sup>174</sup> This provision is put into force through a rulemaking process that the Library of Congress and Copyright Office conducts, which has now resulted in 20 such exceptions ranging from allowances for circumvention to allow a patient to access their own health data<sup>175</sup> to exemptions in support of security research on computer software.<sup>176</sup> These regulations are enacted through a lengthy review process that is repeated every three years, and are only put in place where proponents have demonstrated by a preponderance of the evidence that their uses are non-infringing.<sup>177</sup>

Application of TPM restrictions, while ostensibly aimed at deterring infringement, have serious implications for long-term preservation, access, and research. Many TPMs tie access to proprietary software or platforms, meaning the content becomes unreadable if the platform is discontinued or the TPM scheme is no longer supported. This creates a fragile access environment where e-books and streaming media may effectively vanish when technological dependencies fail. Even simpler forms of digital rights management, such as font encryption or digital watermarking, can obstruct efforts to preserve content in usable formats. When libraries are unable to decrypt fonts or interact with restricted content, the result is preserved data that is inaccessible or unintelligible to researchers. These schemes also impede lawful uses, such as downloading, printing, or remediation for accessibility—uses permitted by courts as a fair use but also uses explicitly permitted under the Copyright Act under Section 121, which allows institutions that have a primary mission to serve the blind or others with disabilities (which many libraries do) to make copies for such users.<sup>178</sup> The barriers TPMs impose are especially

---

owners will hesitate to make their works readily available on the Internet without reasonable assurance that they will be protected against massive piracy. Legislation implementing the treaties provides this protection and creates the legal platform for launching the global digital on-line marketplace for copyrighted works.”)

<sup>173</sup> Whether fair use is a defense to an alleged Section 1201 violation in the absence of falling within one of these regulatory identified categories is a matter of dispute. *Compare Universal City Studios, Inc. v. Corley*, 273 F. 3d 429 (2d Cir. 2001) (rejecting a First Amendment and fair use challenge to Section 1201’s anti-trafficking provisions) with *Chamberlain Group v. Skylink Tech., Inc.*, 381 F. 3d 1178, 1202-03 (Fed. Cir. 2004) (“We conclude that 17 U.S.C. § 1201 prohibits only forms of access that bear a reasonable relationship to the protections that the Copyright Act otherwise affords copyright owners.”). Because of these conflicting precedents, the location of some libraries (e.g., those in the Second Circuit) can make the prospect of liability especially concerning.

<sup>174</sup> 17 U.S.C. § 1201(a)(1)(B) (2025).

<sup>175</sup> 37 C.F.R. § 201.40(7).

<sup>176</sup> 37 C.F.R. § 201.40(18).

<sup>177</sup> 37 C.F.R. § 201.40(b).

<sup>178</sup> See *Authors Guild, Inc. v. HathiTrust*, 755 F. 3d 87, 101 (2d. Cir 2014) (“The [HathiTrust Digital Library] also provides print-disabled patrons with versions of all of the works contained in

acute for users with visual impairments or cognitive disabilities, who often rely on assistive technologies that TPMs block or render unusable.

Recognizing these harms, the exemptions granted by the Library of Congress cover a wide range of academic and library uses, such as making excerpts of films for criticism and comment in documentaries and for teaching purposes,<sup>179</sup> providing accessible copies of works for disabled students,<sup>180</sup> preserving motion pictures,<sup>181</sup> video games,<sup>182</sup> and software<sup>183</sup> by libraries and archives,<sup>184</sup> making copies of motion pictures and literary works for scholarly text and data mining.<sup>185</sup> While several of those exemptions apply only to works embodied in certain physical formats (e.g., covering only motion pictures on DVDs or BluRays), several, such as an exemption for preservation of video games<sup>186</sup> and another for text and data mining research using e-books and motion pictures,<sup>187</sup> specifically allow for circumvention of digital transmissions or licensed, downloaded digital works.<sup>188</sup>

Virtually all of the distributors described here place contractual limits in the EULAs on users' ability to circumvent technological protection measures in any way, including for uses that would otherwise comply with the above listed exemptions. For example, for e-books, Amazon's license states that users, "may not attempt to bypass, modify, defeat, or otherwise circumvent any digital rights management system." Similarly, Apple's license states that users "may not tamper with or circumvent any security technology," Google Play's states users may not "attempt to, or assist, authorize or encourage others to circumvent, disable or defeat any of the security features," and eBooks.com's states that users "promise not to circumvent any measures . . . taken to protect the rights in the eBook." Similarly, streaming services such as Amazon Prime prohibit any "attempt to disable, bypass, modify, defeat, or otherwise circumvent any digital rights management or other content protection system." The Disney license states

---

its digital archive in formats accessible to them. . . . We conclude that this use is also protected by the doctrine of fair use."); 17 U.S.C. § 121 (2025) ("[I]t is not an infringement of copyright for an authorized entity to reproduce or to distribute in the United States copies or phonorecords of a previously published literary work or of a previously published musical work that has been fixed in the form of text or notation if such copies or phonorecords are reproduced or distributed in accessible formats exclusively for use by [blind or print-disabled users.]")

<sup>179</sup> 37 C.F.R. § 201.40(b)(1).

<sup>180</sup> 37 C.F.R. § 201.40(b)(2) & (b)(6).

<sup>181</sup> 37 C.F.R. § 201.40(b)(3).

<sup>182</sup> 37 C.F.R. § 201.40(b)(3).

<sup>183</sup> 37 C.F.R. § 201.40(b)(19).

<sup>184</sup> 37 C.F.R. § 201.40(b)(20).

<sup>185</sup> 37 C.F.R. § 201.40(b)(4)-(5).

<sup>186</sup> 37 C.F.R. § 201.40(b)(19).

<sup>187</sup> 37 C.F.R. § 201.40(b)(4)-(5).

<sup>188</sup> We will not belabor the point since these uses are discussed extensively above, but highlight here some contract terms that are particularly harmful for scholarly work using new computational research techniques. First, the prohibitions noted above—limits on reproduction, and allowing only for personal and non-commercial use—probably foreclose TDM since TDM requires reproduction and is often for research that will be shared broadly, not for personal or entertainment purposes contemplated by EULAs. These anti-circumvention clauses make TDM research yet more difficult.

a user cannot “circumvent or disable any content protection system or digital rights management technology” or “us[e] a robot, spider, scraper or other automated means or manual process.” Likewise, Netflix’s license prohibits “circumvent[ing] . . . content protections” and the use of a spider or scraper or “any data mining, data gathering or extraction method.” Similar restrictions are present in both the HBO and YouTube licenses.

While violation of these EULA anticircumvention terms may not carry the same liability as a violation of the DMCA, which carries potentially hefty statutory damage awards,<sup>189</sup> experience with both researchers and librarians in the field has shown that even liability for breach of contract combined with the potential for loss of access is enough to dissuade many users from pursuing otherwise lawful uses.

### *C. Unwillingness to accommodate library and research interests for consumer market materials*

For the most part, the situation for academic and library users of consumer market materials is bleak. While we do see a few hints from a few licensors that they recognize that lawful uses such as fair use are worth accommodating (for example, the Barnes & Noble Nook EULA which acknowledges fair use), for the most part these EULAs rewrite the rules for how researchers and libraries can use copyrighted materials, creating contractual liability for uses that copyright law allows.

One of the major challenges in this area has been the lack of any sort of commercial interest on the part of the distributors to craft alternative terms that would accommodate academic or library interests. Unlike the agreements discussed above where vendors may have dedicated teams to work with libraries to negotiate terms that can accommodate both commercial and research interests, for these consumer market distributors the prevailing experience is that most librarians and researchers cannot get them to the negotiating table.

When surveying participants at the contract override workshop, we learned that even finding a contact at these large firms who is willing to discuss the matter has been challenging. Enabling more fair uses or other traditionally permissible library and research uses has not been a demonstrated priority of most distributors of popular consumer materials. This, combined with the fact that these firms are increasingly consolidating distribution and offering titles exclusively through their platforms, means that large swaths of culturally important materials are off-limits to researchers and libraries, preventing a wide variety of important research and preservation uses.

### *III. PUBLIC WEBSITES AND TERMS OF SERVICE*

A third area of concern for researchers and libraries comes from the limitations in public websites’ Terms of Service (ToS). These public websites host materials that are freely accessible to the public—often without having to create an account, and encompass materials ranging from those that are uncopyrightable, in the public domain, generated by users (UGC), and assigned or licensed exclusively to the public websites.

---

<sup>189</sup> 17 U.S.C. § 1203(c) (2025).

These materials, especially the UGC and the uncopyrightable data associated with the UGC, are often only available through the specific websites and not accessible anywhere else. Researchers wish to use materials on the public websites to conduct research, and libraries wish to archive certain content from the public websites. For researchers and libraries, these public websites represent an invaluable source of contemporary cultural expressions. Research on online misinformation, for example, would be impossible without collecting and reusing content hosted on public websites. The archiving and preserving of born-digital news and UGC that relate to important political events or natural disasters would also be impossible without automated scraping of content on public websites. These research and archival activities tend to fall under 102(b) and the limitations and exceptions of copyright law.

But in practice, when researchers and libraries access, scrape, download, analyze and/or archive public websites, their ability to do so may be subject to the websites' ToS. The problems presented by ToS are distinct from the consumer contract and negotiated licenses discussed in previous sections, not least because users could be bound by terms they may never see or agree to. As we discuss below in Section IV.C.i., ToS may be more susceptible to challenges on their unenforceability. Still, some courts still treat these unnegotiated—and often unseen—ToS as enforceable contracts simply because a business unilaterally wishes that to be true. This leaves researchers and libraries without recourse when engaging in otherwise lawful research or archival activities, such as scraping non-copyrightable data for research purposes or archiving culturally significant UGC, if such otherwise lawful activities can nevertheless be challenged as a breach of contract.

#### *A. Examples of uses that benefit the public, and ToS that prohibit these uses*

These public website ToS aims to create a digital landscape where public knowledge is fenced off, dictating when the researchers and libraries can document, analyze and preserve our shared online history. The benefits of lawfully reusing content hosted on public websites can be wide ranged: from compiling and communicating real time public health development such as done by the COVID Tracking Project,<sup>190</sup> to providing more transparency to consumers—such as when scrapers archive the pricing history on Ticketmaster or flight booking sites,<sup>191</sup> to allowing better public access to laws such as efforts undertaken by Public.Resources.org,<sup>192</sup> and to allowing continued

<sup>190</sup> The COVID Tracking Project offers the most complete data about COVID-19 in the US. It was cited in more than 1,000 academic papers, including major medical journals.

<sup>191</sup> Even though ToS is what we focus on in this article, it is far from the only legal theory a company can cobble together to stop unwanted uses of their web content. Take flight price tracking for example, Southwest has a track record of sending out copious amount of cease and desist letters citing many potential claims they can pursue. Mike Masnick, *Southwest's Bullshit Lawsuit Over Site That Made \$45 Helping People Book Cheaper Flights*, TECHDIRT, (Jan. 19, 2018), <https://www.techdirt.com/2018/01/19/southwests-bullshit-lawsuit-over-site-that-made-45-helping-people-book-cheaper-flights/>.

<sup>192</sup> [Public.Resources.org](https://www.publicresources.org/) digitizes and publishes the works of the United States Federal Government, including large numbers of court records, US government-produced video, and laws.

access to each administration's websites by the End of Term Web Archive. By contrast, not allowing uses that qualify as fair use (or is otherwise lawful) can harm public interest, such as when Facebook disabled several NYU researchers' accounts to stop them from studying disinformation on the platform.<sup>193</sup> Indeed, the recently proposed mass deletion of online history records—guided by our elected administration—raises the question of whether the right to archive digital materials should in fact extend far beyond current norms, lest we risk erasing entire generations' cultural and historical record.<sup>194</sup>

When we examine public websites' ToS, we notice that potential socially beneficial activities, such as those listed in the previous paragraph, are almost uniformly not allowed. LinkedIn's ToS<sup>195</sup> states that anyone who "create an account or access or otherwise use any of [LinkedIn] Services" is bound by their Terms, which include the following section which clearly extends LinkedIn's control of downstream use well beyond the traditional contours of copyright, and broadly restricts activities that would be socially beneficial or at least harmless to LinkedIn:

You agree that you will not:

...

2. **Develop**, support or use software, devices, scripts, robots or any other **means or processes (such as crawlers**, browser plugins and add-ons or any other technology) **to scrape or copy the Services, including profiles and other data** from the Services;

<sup>193</sup> EFF discussed this issue in more detail, documenting how this incident is but one of many examples of Facebook's problematic policy stance. Rory Mir & Cory Doctorow, *Facebook's Attack on Research Is Everyone's Problem*, ELEC. FRONTIER FOUND. (Aug. 12, 2021), <https://www.eff.org/deeplinks/2021/08/facebooks-attack-research-everyones-problem>.

<sup>194</sup> Seemingly stable online records can be more precarious than we think. If we want to keep a consistent and complete record of our shared history, allowing and even encouraging citizens to scrape and archive may be the only way forward. See, e.g., Tara Copp, Lolita C. Baldor & Kevin Vineys, *War Heroes and Military Firsts Are Among 26,000 Images Flagged for Removal in Pentagon's DEI Purge*, ASSOCIATED PRESS (Mar. 7, 2025), <https://apnews.com/article/dei-purge-images-pentagon-diversity-women-black-8efcfaec909954f4a24bad0d49c7807>; Martin Kaste, *Trump Took Down Police Misconduct Database, but States Can Still Share Background Check Info*, NPR (Feb. 28, 2025), <https://www.npr.org/2025/02/28/nx-s1-5305281/trump-police-misconduct-database-background-checks>. Of course, the problem is not better in the private sector. Glyn Moody, *We Risk Losing Access to the World's Academic Knowledge, and Copyright Makes Things Worse*, WALLED CULTURE (Mar. 27, 2024), <https://walledculture.org/we-risk-losing-access-to-the-worlds-academic-knowledge-and-copyright-makes-things-worse/>.

<sup>195</sup> *User Agreement*, LINKEDIN (Nov. 20, 2024), <https://www.linkedin.com/legal/user-agreement>, <https://www.linkedin.com/legal/user-agreement> (last accessed June 13, 2025). If the person reading these terms does not need to research what any of the terms mean nor desire to click on any links in the documents to find out more, it would take an average reader (reading at a college level) over 40 minutes to go over LinkedIn's ToS and Privacy Policy. Of course, "not requiring additional research to understand the terms" is an unrealistic assumption, given many of the activities discussed in the Terms are commonly undertaken by users who do not know the term-of-art describing such activities (e.g., deep-linking, overlaying, etc.).

3. Override any security feature or bypass or circumvent any access controls or use limits of the Services (such as search results, profiles, or videos);

4. **Copy, use, display or distribute any information (including content) obtained from the Services**, whether directly or through third parties (such as search tools or data aggregators or brokers), without the consent of the content owner (such as LinkedIn for content it owns);

...

9. Reverse engineer, decompile, disassemble, decipher or otherwise attempt to derive the source code for the Services or any related technology that is not open source;

...

12. **Deep-link** to our Services for any purpose other than to promote your profile or a Group on our Services, without LinkedIn's consent;

13. Use bots or other unauthorized automated methods to access the Services, add or download contacts, send or redirect messages, create, comment on, like, share, or re-share posts, or otherwise drive inauthentic engagement;

...

14. Overlay or otherwise modify the Services or their appearance (such as by inserting elements into the Services or removing, covering, or obscuring an advertisement included on the Services);

...

LinkedIn's ToS exemplifies how public-facing platforms can assert expansive, contract-based control over data and content that is otherwise publicly visible and often legally reusable under copyright law. By broadly prohibiting scraping, automated access, deep-linking, and even redistributing publicly viewable profile information, LinkedIn's ToS seek to restrict activities that serve vital public interests—such as research projects that study labor market trends and employment equity issues. These sweeping contractual restrictions far exceed the limited rights granted to copyright holders under Section 106. Compounding the issue, LinkedIn is not even the copyright owner of much of the UGC it attempts to control. The ToS attempt to create quasi-intellectual property rights that bind the public.

LinkedIn's example is not singular in any way. While organized and formulated in a very different way, the ToS of Reddit<sup>196</sup> nevertheless touches on similar prohibitions across different sections. If you would like “a personal, non-transferable, non-exclusive, revocable, limited license to ... access and use the [Reddit] Services,” you must promise “your complete and ongoing compliance with [Reddit] Terms,” including the following:

### 3. Your Use of the Services

...

...you may not...:

- license, sell, transfer, assign, **distribute, host**, or otherwise commercially exploit the Services or **Content**;
- **modify, prepare derivative works of**, disassemble, decompile, or reverse engineer any part of the Services or **Content**; or

<sup>196</sup> User Agreement, REDDIT (May 29, 2025), <https://redditinc.com/policies/user-agreement> (last accessed June 13, 2025). Reddit's ToS and Privacy Policy similarly takes over 40 minutes to read—again, assuming you fully comprehend all the relevant terms without requiring additional research.

- access the Services or Content in order to build a similar or competitive website, product, or service, except as permitted under any Additional Terms (as defined below).

...

#### 5. Your Content

...

When Your Content (defined earlier as “Content created with or submitted to the Services by you or through your Account”) is created with or submitted to the Services, you grant us a worldwide, royalty-free, perpetual, irrevocable, non-exclusive, transferable, and **sublicensable license to use, copy, modify, adapt, prepare derivative works of, distribute, store, perform, and display Your Content** and any name, username, voice, or likeness provided in connection with Your Content in all media formats and channels now known or later developed anywhere in the world.

...

#### 7. Things You Cannot Do

...you may not...:

...

**Access, search, or collect data** from the Services by any means (automated or otherwise) except as permitted in these Terms or in a separate agreement with Reddit (we conditionally grant permission to crawl the Services in accordance with the parameters set forth in our robots.txt file, but scraping the Services without Reddit’s prior written consent is prohibited);

Just like LinkedIn’s ToS, Reddit’s ToS attempt to create broad para-IP right, substituting private, non-negotiated contractual control which favors one party for the traditional boundaries of copyright law that balances private and public interest. The ToS allow Reddit to claim copyright-like powers over content it does not own and impose constraints that stifle academic research and limit the preservation of digital culture.

LinkedIn and Reddit websites are mostly accessible to the public even without creating accounts on those sites. Some public websites, however, require users to create an account and click agree to the ToS before they may proceed to engage with content on the websites. These websites predominantly host UGC, which often include materials that are uncopyrightable, in the public domain, or otherwise lawful for reuse. In any case, these public websites do not purport to be the rightsholders of the UGC, but they nevertheless exert broad control over the UGC, as well as any otherwise lawful downstream reuse of the UGC.

[X.com](https://x.com/en/tos) is often only viewable to registered users. X Corp’s ToS<sup>197</sup> start out by saying “These Terms of Service (“Terms”) are part of the User Agreement – a legally binding contract governing your use of X. You should read these Terms of Service (“Terms”) in full.” The Terms state:

If you want to **reproduce, modify, create derivative works**, distribute, sell, transfer, publicly display, publicly perform, transmit, or otherwise use the Services or Content on the Services, you must use the interfaces and instructions we provide, except as permitted through the Services, these Terms, or the terms provided on <https://developer.x.com/developer-terms>. **Otherwise, all such actions are strictly**

<sup>197</sup> *Terms of Service*, X, (Nov. 15, 2024), <https://x.com/en/tos>.



**prohibited.** If you are a security researcher, you are required to comply with the rules of our Vulnerability Reporting Program (<https://hackerone.com/x>).<sup>198</sup>

[...]

You also agree not to misuse the Services, for example, by interfering with them or **accessing them using a method other than the interface** and the instructions that we provide. You agree that you will not work around any technical limitations in the software provided to you as part of the Services, or reverse engineer, decompile or disassemble the software, except and only to the extent that applicable law expressly permits. You may not do any of the following while accessing or using the Services: (i) access, tamper with, or use non-public areas of the Services, our computer systems, or the technical delivery systems of our providers; (ii) probe, scan, or test the vulnerability of any system or network or breach or circumvent any security or authentication measures; (iii) **access or search or attempt to access or search the Services by any means (automated or otherwise) other than through our currently available, published interfaces that are provided by us** (and only pursuant to the applicable terms and conditions), unless you have been specifically allowed to do so in a separate agreement with us (NOTE: crawling or scraping the Services in any form, for any purpose without our prior written consent is expressly prohibited) [...] (*the block of text goes on*)

As with LinkedIn and Reddit, these restrictions listed in X's ToS arise not from public legislation or court rulings, but from unilaterally drafted, non-negotiable, private contract terms. These examples include terms that are plainly unenforceable. For example, provisions that prohibit any modification or redistribution of user-generated content would, if enforced literally, produce absurd results: it burdens free speech in a way that is repugnant to basic constitutional values. If websites truly intended to enforce these terms as written, entire fields like social media research would cease to exist. Yet despite the clear overreach of some terms, they are presented alongside enforceable provisions as if all have equal legal weight—because there is no harm to the drafter of the ToS if their language is overbroad. There is little incentive to draft only legally permissible terms when courts routinely step in to sever the unenforceable portions and enforce the rest, effectively rewarding aggressive drafting strategies.<sup>199</sup> Professor Mark

<sup>198</sup> See X, *supra* note 198. What is permitted for the users is in stark contrast with the broad permission X purports to receive from its users: "By submitting, posting or displaying Content on or through the Services, you grant us a worldwide, non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display, upload, download, and distribute such Content in any and all media or distribution methods now known or later developed, for any purpose. [...] You agree that this license includes ... use with and training of our machine learning and artificial intelligence models, whether generative or another type. [...] Such additional uses by us, or other companies, organizations or individuals, is made with no compensation paid to you. [...] You represent and warrant that you have, or have obtained, all rights, licenses, consents, permissions, power and/or authority necessary to grant the rights granted herein for any Content that you submit, post or display."

<sup>199</sup> Charles A. Sullivan, *The Puzzling Persistence of Unenforceable Contract Terms*, 70 OHIO ST. L. J. 1127 (2009). For more examples of intentionally unenforceable clauses, see Cheryl B. Preston, "Please Note: You Have Waived Everything": Can Notice Redeem Online Contracts? 3 AM. U. L. REV. 555 (2015).

Lemley has explained the absurdity of this unfettered corporate power with this apt analogy: “The reason my “no-trespassing” sign is effective in the real world is not because there is any sort of agreement to abide by it, but because the law already protects my land against intrusion by another. If the sign read “no walking on the road outside my property,” no one would think of it as an enforceable agreement.”<sup>200</sup>

Worse still, as we explain in Section V, most courts—regardless of how a given case is eventually resolved—will at least entertain the possibility that some overly-burdensome terms are nevertheless enforceable. Take *Internet Archive v. Suzanne Shell*,<sup>201</sup> for example. The breach of contract claim was allowed to proceed despite the website ToS prescribing exorbitant fees for copying the website. According to the court:

These terms include charging the user \$5,000 for each individual page copied “in advance of printing,” granting Shell a perfected security interest of \$250,000 “per each occurrence of unauthorized use” of the website in all of the user’s land, assets and personal property, the user agreeing to pay “\$50,000 per each occurrence of failure to pre-pay” for use of the website, “plus costs and triple damages,” and agreeing to waive numerous defenses in any claims by Shell against the user.

Recognizing that “[t]hese requirements and damages lie well beyond the protections Shell receives through the Copyright Act,” the court nevertheless refused to dismiss the breach of contract claim, and the parties eventually settled. Internet Archive’s Wayback Machine—and the vast public benefit it provides—could not be sustained, if courts imposed a duty to parse individualized ToS for every archived page. From its launch in 2001, the Wayback Machine has supported vital historical, academic, legal, and journalistic work. It preserves content that would otherwise be lost, enables researchers to study web development over time, and even supports civil litigation (such as in patent cases) and democratic accountability. Wikipedia editors use it extensively to prevent link rot; scholars have used it to study how website content affects corporate growth; journalists rely on it to expose misinformation and verify facts; and courts have admitted its archived pages as evidence. All this would be made impossible if courts allow noncommercial archival and informational uses to be gravely burdened by idiosyncratic ToS. The case of Wayback Machine, and other socially beneficial research and archival use cases of public websites, drives home the warning by Friedrich Kessler “that freedom of contract must mean different things for different types of contracts. Its meaning must change with the social importance of the type of contract and with the degree of monopoly enjoyed by the author of the standardized contract.”<sup>202</sup>

Beyond having to scrutinize the legal enforceability of these terms, there are at least five immediate challenges for researchers and libraries working in the U.S. One, it is nearly impossible to keep track of multiple lengthy, differently-structured ToS that are at the same time subject to change at any moment based on the whims of the Service provider. Two, for a prudent and risk-averse user, the lack of clarity and consistency in ToS discourages research use, archival use, interoperability, private customization, and

<sup>200</sup> Mark A. Lemley, *Terms of Use*, 91 MINN. L. REV. 459 (2006).

<sup>201</sup> 505 F.Supp.2d 755 (D. Colo. 2007).

<sup>202</sup> Friedrich Kessler, *Contracts Of Adhesion—Some Thoughts About Freedom Of Contract*, 43 COLUM. L. REV. 629 (1943).

fair use and use of public domain content in general, that are not explicitly permitted under the website terms. Three, dominant platforms are often essential gateways to public discourse, and there's no meaningful alternative for the general public, researchers, or libraries and archives. Four, U.S. researchers and libraries are falling behind due to a lack of clear protection for their lawful uses: jurisdictions like the E.U. are taking more aggressive steps to safeguard user rights, preserve access to knowledge, and regulate the market power of large tech intermediaries. Finally, the cost related to litigation is prohibitive, often a greater deterrent for researchers and libraries than any actual damages the plaintiffs can eventually prove.

### *B. The legality of the prohibitive ToS*

As a general matter, ToS requires some indication of assent to be enforceable. Though what counts as "assent" is not entirely clear, courts are increasingly seen to engage in fact-specific analysis instead of assuming the presence of ToS alone proves its legal validity, especially when there is no consent mechanism (e.g., click wrap). So far, one circuit court has held that allowing private corporations to decide who can collect and reuse their public data—data generated by their users no less—"risks the possible creation of information monopolies that would disserve the public interest."<sup>203</sup> Section V.C.i. discusses contract validity in more detail.

Beyond scrutinizing over general contract formation, when it comes to how copyright laws may interact with ToS, U.S. jurisprudence is still underdeveloped, with very few cases—and no legislation—directly addressing if website ToS can lawfully prevent the public from accessing the webpages to make fair uses or even from using public domain or uncopyrightable materials.

Many arguments have been advanced by website owners to establish the legality and enforceability of these ToS. But their assertions are dubious, especially when the website owners are not the copyright owners of the digital content being accessed or scraped.

The ambiguity and constraint created by ToS is exacerbated by technological developments. In *Consent in Crisis: The Rapid Decline of the AI Data Commons*, the authors explained how signals of user consent on the web—particularly as expressed through ToS and robots.txt files—are increasingly incoherent and difficult to interpret. The authors conducted a large-scale, longitudinal audit of over 14,000 domains used in AI training datasets, finding that nearly half of all web tokens in key corpora (e.g., C4, RefinedWeb) are now subject to some form of restriction on downstream uses such as AI training and crawling. Yet these restrictions are often expressed in conflicting ways: robots.txt files may permit access while ToS prohibit it—or vice versa—and many restrictions rely on outdated user-agent names. The paper shows that existing consent signaling protocols were not designed for the current landscape and fail to meaningfully capture content creators' intentions. Rather than a system of informed, enforceable user

---

<sup>203</sup> *hiQ Labs, Inc. v. LinkedIn Corp.*, 31 F. 4th 1180 (9th Cir. 2022). This quote was relied on in *X Corp. v. Bright Data Ltd.*, 733 F.Supp.3d 832 (N.D. Cal. 2024), before the court held that X cannot stop Bright Data from scraping public facing tweets.

choice, we see a fragmented patchwork that disproportionately affects public-interest uses such as archiving, academic research, and nonprofit development.<sup>204</sup>

*C. Recent cases, failed defenses, and undesirable consequences*

Notable recent litigations in this area include *MLGenius Holdings LLC v. Google LLC*<sup>205</sup>, *hiQ Labs, Inc. v. LinkedIn Corp.*<sup>206</sup>, *Meta Platforms, Inc. v. Bright Data Ltd.*<sup>207</sup>, *X Corp. v. Center for Countering Digital Hate*<sup>208</sup>, and *X Corp. v. Bright Data Ltd.*<sup>209</sup>, resulting in some mixed rulings for web scraping and the reuse of public websites in general. We see in these cases how courts grapple with the balance between private control and public interest.

Many consider *hiQ Labs, Inc. v. LinkedIn Corp.* and *X Corp. v. Bright Data Ltd.* to be the most important cases in web scraping jurisprudence. The *hiQ Labs* case centered on whether LinkedIn could block hiQ Labs from scraping public profile data, with the Ninth Circuit initially siding with hiQ by affirming that scraping public data likely did not violate the CFAA (Computer Fraud and Abuse Act). However, in a later ruling, the district court granted LinkedIn summary judgment on its breach of contract claim, finding that hiQ clearly violated LinkedIn's User Agreement by scraping data and using fake accounts. In *Bright Data*, by contrast, the Northern District of California held that X's Terms of Service were unenforceable under the doctrine of conflict preemption, because enforcing them would undermine federal copyright policy—this is a rather new development that we haven't seen in other courts in recent years. It is also unclear at this point, as the case is still active, whether X Corp can overcome this hurdle by showing it has sustained actual harm in its servers due to the web-scraping activities. Courts struggle to define the boundaries of federal copyright policy that remain off-limits to private contractual override, and to articulate when and how those public policy concerns must factor into contract enforcement. We discuss preemption in section V.C.ii.

Some failed defenses web-scrapers have attempted in such breach of ToS litigations include:

**Ambiguous Terms.** In *hiQ Labs, Inc. v. LinkedIn Corp.*, hiQ pointed out that LinkedIn's User Agreement warns users that public data may be copied and used by others.<sup>210</sup> The court found the warning did not equate a “blessing” for scrapers to take and use public-facing data freely, and proceeded to conclude that “the relevant language of the User Agreement unambiguously prohibits hiQ's scraping and unauthorized use of the scraped data.”

**Independent Contractors.** In *hiQ Labs, Inc. v. LinkedIn Corp.*<sup>211</sup>, hiQ tried to argue that it was not responsible for the conduct of its contractors. The court was not

<sup>204</sup> Shayne Longpre et al., *Consent in Crisis: The Rapid Decline of the AI Data Commons* (July 20, 2024), <https://arxiv.org/abs/2407.14933>.

<sup>205</sup> 2022 U.S. App. LEXIS 6206 (2d Cir. 2022).

<sup>206</sup> 639 F. Supp. 3d 944 (N.D. Cal. 2022).

<sup>207</sup> 2024 WL 251406 (N.D. Cal. 2024).

<sup>208</sup> 724 F. Supp. 3d 948 (N.D. Cal. 2024).

<sup>209</sup> 733 F.Supp.3d 832 (N.D. Cal. 2024).

<sup>210</sup> 639 F. Supp. 3d 944 (N.D. Cal. 2022).

<sup>211</sup> *Id.*

persuaded, and instead found the independent contractors to be hiQ's agents as they created fake user accounts to access LinkedIn data.

**Unfair Competition.** In *hiQ Labs, Inc. v. LinkedIn Corp.*,<sup>212</sup> hiQ claimed it suffered damages caused by LinkedIn's cease-and-desist letter. HiQ asked the court to apply Noerr-Pennington doctrine's "sham" analysis and find that the "[cease-and-desist] letter was merely the 'enforcement mechanism . . . [of LinkedIn's] anticompetitive scheme.'" The court concluded that the California litigation privilege is broader and did not have a "sham" exception: as long as litigation was actually contemplated seriously—whatever the underlying motivation may be, prelitigation communications cannot be held to be illegal.

**Unconscionability.** Generally speaking, a contract needs to be both procedurally and substantively unconscionable for a court to refuse to enforce it. In *Internet Archive v. Suzanne Shell*, the court refused to consider the Internet Archive's unconscionability defense, stating that the Internet Archive failed to show unequal bargaining power, its absence of meaningful choice, or why the ToS are overly favorable to the website owner. In *hiQ Labs, Inc. v. LinkedIn Corp.*,<sup>213</sup> hiQ tried arguing unconscionability based on the fact that each visitor of the website was in violation of LinkedIn's ToS when they requested data browsing public profiles—in a way indistinguishable from scrapers, so that LinkedIn could selectively enforce its ToS; but the court concluded that no reasonable user would interpret the ToS in such a way. HiQ likely failed because it couldn't point to any unconscionability related to the ToS's prohibition of scraping and the use of scraped data; without such substantive unconscionability, the court did not address procedural unconscionability.

These defenses will likely continue to fail unless given very unique case facts. The public websites' ToS are more unilateral than the take-it-or-leave-it licensing terms described in the sections above. These one-sided ToS threaten to divest our rights to study and preserve public information. They often run contrary to the goals of our copyright laws and restrict what otherwise should be freely used. Without intervention, the internet risks becoming an ephemeral space controlled by private entities, rather than a lasting and accessible record of shared knowledge.

#### IV. SOLUTIONS

At the workshop precipitating this paper, copyright and library experts discussed a number of potential ways to address the above contractual override concerns, many of which we attempt to summarize in this section. Which of these pathways should libraries and scholars actually pursue? The authors' preference is for a fulsome solution: comprehensive but tailored Federal legislation ensuring that copyright exceptions and limitations for researchers and libraries may not be overridden by contracts. The European Union has implemented this effect with respect to multiple limitations and exceptions, and this framework is both our goal and gold standard for addressing contractual override of research and library rights. We discuss the benefits and hurdles of this inclusive approach below.

---

<sup>212</sup> *Id.*

<sup>213</sup> *Id.*

Yet, we also believe it is important also to consider the variety of other options discussed at the forum that might help abate, albeit less panoramically, contractual override problems for scholarly research and libraries. Pursuing these more limited solutions may be akin to plastering modest patches over leaks that have sprung in the copyright limitations and exceptions framework; but they may be important (if only temporary) patches nonetheless. Loosely speaking, these patches may be classified within three broad categories: (1) legislative and regulatory; (2) private ordering; and (3) judicial. We account for them in the context of the three key contract types discussed above: library-licensed materials, consumer EULA materials, and web-based materials used in research—we evaluate them under the lens of their ability to provide only partial relief.

#### *A. Legislative and Regulatory Approaches*

##### 1. Fulsome federal solution: the European Union framework

To a great extent, the European Union has already achieved our end goal through the Copyright in the Digital Single Market Directive, Article 7. Article 7 provides that “Any contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 shall be unenforceable.” In turn, Article 3 covers the text and data mining exception for scientific research by research organizations and cultural heritage institutions; Article 5 addresses the use of copyright-protected digital works in non-commercial instruction; and Article 6 ensures the right of cultural heritage institutions to reproduce any materials in their permanent collections for preservation purposes. Thus, under CDSM Article 7, none of those rights espoused in Articles 3, 5, and 6 may be taken away by any agreement.

In the U.S. a nominally-similar Federal approach to override has been considered by the U.S. Congress once before though only in the context of “non-negotiable” consumer contracts—and notably at a time before the proliferation of digital licensing problems discussed above. Through H.R. 5522 (introduced in the 107th Congress in 2002), Representative Zoe Lofgren sought to “Protect[] Fair Use and Consumer Expectations” by adding a new section of the Copyright Act providing:

(b) Effect of Licenses. When a digital work is distributed to the public **subject to non-negotiable license terms**, such terms shall not be enforceable under the common laws or statutes of any State **to the extent that they restrict or limit any of the limitations on exclusive rights under this title** (emphasis added).<sup>214</sup>

The bill was referred to committee with no further action taken, and Rep. Lofgren introduced it again under a different title in 2003 but with similar results.<sup>215</sup> We are aware of no further Congressional efforts to pursue a similar plan.

In our view, Congress should examine a “CDSM Article 7” type of approach, but with more precise effect than the Lofgren bills. It is likely that those 2002-03 bills

<sup>214</sup> H.R. 5522, 107th Cong. (2002).

<sup>215</sup> H.R.1066, 108th Cong. (2003).

suffered simultaneously from overbreadth (in their rather impractical attempt to eliminate override for all non-negotiable licenses) and lack of focus (in applying to consumer agreements and failing to limit themselves to less controversial contracts governing just research and education).<sup>216</sup> Both shortcomings could be resolved through greater specificity with regard to pertaining only to research and libraries. Indeed, this type of legal instrument(s) is also encouraged by the Work Program of the World Intellectual Property Organization's Standing Committee on Copyright and Related Rights (SCCR), which since 2012 has had a mandate to address the issue of rights and exceptions for research and education, library preservation, and more.<sup>217</sup> The SCCR Work Program now includes consideration of legal instruments to address the overridability of these limitations and exceptions.<sup>218</sup> In part to inform the next steps within the SCCR Work Program, the Global Network on Copyright User Rights held a symposium on protecting user rights from contract override in 2023,<sup>219</sup> and in 2025 deliberated over principles for copyright reform in the digital environment, including as to contractual provisions that otherwise would impede the effective exercise or enjoyment of exceptions and limitations.<sup>220</sup>

While the EU's current CDSM and AI Act framework aims to preclude override for multiple limitations and exceptions simultaneously, the CDSM still bears some shortcomings in implementation—as evidenced by examining how CDSM's Article 7 (no override) and Article 3 (TDM permitted) interact in practice. To understand this interaction, it is helpful to recognize first that as of June 2025, more than forty countries (including all those within the EU<sup>221</sup>) expressly protect text mining rights for scientific research.<sup>222</sup> The EU's recent AI Act extends the Article 7 and Article 3 rights to the use of AI when scholars at research institutions train AI with copyrighted works in the course of their research.<sup>223</sup> Copyright owners may not opt out of having their works used for TDM or for training AI in the course of that TDM performed in accordance with Article 3; the only thing Article 3 builds in is the opportunity for copyright owners to

<sup>216</sup> *Id.*; H.R. 5522, 107th Cong. (2002).

<sup>217</sup> Sean Flynn, *A Step Forward: World Intellectual Property Organization's Copyright Committee Inches Forward on Broadcast and Limitations*, INFO JUSTICE (Apr. 21, 2025), <https://infojustice.org/archives/46253>; World Intellectual Prop. Org., Standing Comm. on Copyright & Related Rights, Proposal by Chile on the Subject "Exceptions and Limitations to Copyright and Related Rights," WIPO Doc. SCCR/12/3 (Nov. 2, 2004), [https://www.wipo.int/meetings/en/doc\\_details.jsp?doc\\_id=34747](https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=34747)

<sup>218</sup> World Intellectual Prop. Org., Standing Comm. on Copyright & Related Rights, *Work Program on Exceptions and Limitations Adopted by the Committee*, WIPO Doc. SCCR/43/8 REV. (Mar. 17, 2023), [https://www.wipo.int/edocs/mdocs/copyright/en/scer\\_43/scer\\_43\\_8\\_rev.pdf](https://www.wipo.int/edocs/mdocs/copyright/en/scer_43/scer_43_8_rev.pdf)

<sup>219</sup> User Rights Network Symposium: Protecting Copyright User Rights from Contractual Override, AM. U. WASH. COLL. L., <https://www.american.edu/wcl/impact/initiatives-programs/pijip/events/user-rights-network-symposium-protecting-copyright-user-rights-from-contractual-override.cfm> (last visited July 1, 2025).

<sup>220</sup> User Rights Symposium 2025, AM. U. WASH. COLL. L., <https://www.american.edu/wcl/impact/initiatives-programs/pijip/events/user-rights-symposium-2025.cfm> (last visited July 1, 2025).

<sup>221</sup> Council Directive 2019/790, 2019 O.J. (L 130) 92 (EU).

<sup>222</sup> Band, *supra* note 43.

<sup>223</sup> Parliament and Council Regulation 2024/1689, 2024 O.J. (L 1689) 1 (EU).



require reasonable security measures.<sup>224</sup> Publishers are thus obligated to, and do, preserve fair use-equivalent research exceptions for TDM and AI for research organizations within the EU.<sup>225</sup>

Nevertheless, publishers engaging with academic libraries in the EU still endeavor to create either uncertainty for scholars around what is permitted, or impose a “choice of law” provision that would apply some foreign law that does afford those same protections.<sup>226</sup> Communia, an international association that “advocates for policies that expand the Public Domain and increase access to and reuse of culture and knowledge,” recently issued a report based on interviews with licensing managers from public and academic libraries across Europe.<sup>227</sup> Communia found that publishers endeavored to ban AI tools for research uses even though protected under the national TDM exceptions that implement Article 3 of the CDSM Directive—creating legal uncertainty for researchers.<sup>228</sup> Additionally, publishers introduced foreign law on the construction and application of rights under the agreements which “may be enough to circumvent” the prohibition on TDM and AI override imposed by Article 7 of the CDSM Directive.<sup>229</sup>

Learning what we can from the experience of libraries and scholars in the EU, we believe the US Congress could improve on the EU framework by enacting a statute similar to Article 7 of the EU’s CDSM Directive, yet also addressing or penalizing some publishers’ or websites’ efforts (as Communia found) to work around the exception in research contexts. Will we be successful in achieving such a model? We are uncertain. The reality is that the prospects of Federal legislative reforms are challenging. The authors of this paper have participated in preliminary outreach to Congressional staffers on the issue of overriding computational research rights, with scant support.<sup>230</sup> Moreover, there may be little appetite from the higher education community to engage in this debate amidst substantial other cuts to research funding.<sup>231</sup> Further, as the Association of Research Libraries explained, “any discussion about amending the

<sup>224</sup> Directive 2019/790 (Article 3: “Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.”)

<sup>225</sup> *Id.* See also COMMUNIA, *supra* note 14.

<sup>226</sup> COMMUNIA, *supra* note 14.

<sup>227</sup> *Id.*

<sup>228</sup> *Id.*

<sup>229</sup> *Id.*

<sup>230</sup> Katherine Klosek, *Protecting AI Usage and Training Rights: Three Ideas for Congress*, ASS’N OF RSCH. LIBRARY., (Apr. 18, 2024), <https://www.arl.org/blog/protecting-ai-usage-and-training-rights-three-ideas-for-congress/>.

<sup>231</sup> Research funding cuts have preoccupied discussions within American universities. See, e.g. Cuts to Federal Science Spending Will Cost Every American, UNIV. CAL., <https://www.universityofcalifornia.edu/news/cuts-federal-science-spending-will-cost-every-american> (last visited July 2, 2025); Reports Say Cuts to Federal Research Funding Would Undermine Economic Growth, AAU, <https://www.aau.edu/newsroom/leading-research-universities-report/reports-say-cuts-federal-research-funding-would> (last visited July 2, 2025).

Copyright Act would certainly get the attention of copyright owners and their lobbyists, and may result in unintended consequences that are worse than the status quo.”<sup>232</sup>

## 2. Section 108 reform for preservation and interlibrary lending

At the February 2025 workshop that we held, some scholars raised the possibility of targeting override of libraries’ preservation and interlibrary lending rights through reform of Section 108. Section 108(f)(4) provides that “any contractual obligation assumed by a library or archives upon obtaining a work for its collections supersedes the institution’s privileges under section 108.”<sup>233</sup> As a result, libraries can only infrequently rely on Section 108 to safeguard preservation or interlibrary loan rights, given that commercial contracts do typically override Section 108 in the ways set forth above. Accordingly, legislative reform efforts could seek to eliminate the override language in 108(f)(4). We believe the upshot of doing so would be of little value, however, unless Fair Use is *also* protected from contractual override. That is because subsection (f)(4) already provides that Section 108 in no way “affects the right of fair use as provided by section 107”—meaning Fair Use already remains in libraries’ arsenal for preservation and interlibrary loan purposes, provided libraries can actually negotiate to preserve Fair Use.<sup>234</sup> And as demonstrated *supra*, it is the contracts that take away Fair Use rights that have the outsized impact on the breadth and depth of work that libraries can do.

Section 108 reform has been considered in the past, and through those efforts scant consensus has been reached about revising 108(f)(4) to address override. The first efforts to address override and Section 108 came in 2005, when the Copyright Office charged a “study group” to address Section 108’s viability in the midst of digital technologies that were “radically transforming how copyrighted works are created and disseminated.”<sup>235</sup> The Study Group issued a report in 2008 indicating deep division over “whether section 108, especially the preservation and replacement exceptions, should trump contrary terms in non-negotiable agreements.”<sup>236</sup>

Congress revisited the matter in 2016, with the Copyright Office issuing a Notice of Inquiry that specifically sought input on “how privileging some of the section 108 exceptions over conflicting contractual terms would affect business relationships between rights-holders and libraries, archives, and museums.”<sup>237</sup>

In 2017, the Copyright Office issued the resulting “discussion document,” in which they put forward a compromise approach that would have preserved contract override of Section 108 except in circumstances of “non-negotiable contract terms”—such as “click-wrap” licenses.<sup>238</sup> For these non-negotiable contracts, the Copyright Office

<sup>232</sup> Klosek, *supra* note 108, at 10.

<sup>233</sup> 17 U.S.C. § 108(f)(4).

<sup>234</sup> *Id.*

<sup>235</sup> Section 108 Study Group, The Section 108 Study Group Report (2008), <https://www.section108.gov/docs/Sec108StudyGroupReport.pdf>.

<sup>236</sup> *Id.* at p. 120.

<sup>237</sup> Section 108: Draft Revision of the Library and Archives Exceptions in U.S. Copyright Law, 81 Fed. Reg. 36,594, 36,599 (June 7, 2016).

<sup>238</sup> U.S. COPYRIGHT OFFICE, SECTION 108 OF TITLE 17: A DISCUSSION DOCUMENT OF THE REGISTER OF COPYRIGHTS, (2017), <https://www.copyright.gov/policy/section108/discussion-document.pdf>.

essentially proposed eliminating the availability of infringement claims for Section 108 violations, but leaving intact a copyright owner's ability to file a breach of contract claim. The discussion paper did not result in any statutory changes.

The library community may not have welcomed statutory changes in any event. Although some libraries recognized the surface-level logic of reforming Section 108 to better protect library preservation and loan rights, other libraries expressed the view that the legislative reopening of Section 108 (absent broader reform) poses credible risk that copyright owners would seek to eliminate the Fair Use reservation clause in subsection (f)(4).<sup>239</sup> The Association of Research Libraries has asserted that, "there is no dispute that Section 108, originally written for the 1976 Copyright Act, could benefit from some updates for the digital age. However, these revisions are not necessary because Fair Use sufficiently updates Section 108 to cover activities necessary in the digital age."<sup>240</sup> As we believe Section 107 offers far greater utility for libraries' ability to rely upon copyright exceptions to begin with, a better investment of resources—and one that would not threaten the removal of Section 107's protections from within Section 108—would be to address contract override more holistically through another mechanism discussed herein.

### 3. Federal anti-SLAPP to disincentivize threats against libraries

Experts at the February 2025 workshop also raised a rather niche Federal legislative possibility, and one altogether unlikely to gather momentum: a federal anti-SLAPP (strategic lawsuits against public policy) law. Anti-SLAPP statutes protect defendants by providing procedural mechanisms to quickly dismiss meritless lawsuits that target constitutionally protected speech, shifting the burden to plaintiffs to demonstrate a probability of prevailing on the merits while allowing defendants to recover attorney's fees when they successfully invoke the statute's protections. Arguably, such a law could reduce the chilling effect of publishers' threats on academic institutions that wish to challenge the enforceability of restrictive contractual terms, or protect institutions from being put in a position of having to defend themselves against unfounded breach claims.

As such, a Federal anti-SLAPP law could help provide: (1) an early dismissal mechanism for contract claims when institutions are exercising Fair Use rights, and/or publishers are retaliating against exercise of Fair Use rights; (2) fee shifting protection requiring plaintiffs (largely publishers) to pay attorneys' fees when cases are dismissed—dissuading publishers from filing breach claims when the acts at issue are likely Fair Use; and (3) stays of discovery following the filing of lawsuits that are retaliatory against Fair Uses, again limiting publishers' incentives to use costly discovery as a way to compel early settlement. Theoretically, this approach restores some balance between large commercial publishers and underfunded academic institutions. It could also support academic freedom by empowering researchers to use copyrighted materials in ways that may be contractually restricted but permissible under

<sup>239</sup> Krista Cox, Thoughts on the Section 108 Reform Discussion Draft, ASS'N OF RSCH. LIBRARY. (Sept. 19, 2017), <https://www.arl.org/blog/thoughts-on-the-section-108-reform-discussion-draft/>.

<sup>240</sup> *Id.*

Fair Use. Largely, however, we address it only for the sake of fullness in our consideration of options, a “nice to have” bandage for a much larger problem.

#### 4. State legislative solutions

State legislative solutions—at least with respect to *some* aspects of library e-book licensing, as we explain below—may hold promise. In May, 2025 Connecticut enacted Public Act No. 25-9, titled “An Act Prohibiting Libraries From Agreeing To Certain Terms In Electronic Book And Digital Audiobook License Agreements Or Contracts.”<sup>241</sup> The law addresses restrictive e-book licensing practices that interfere with some of libraries’ traditional functions.<sup>242</sup> The legislation prohibits libraries from entering into contracts with publishers that simultaneously restrict both the number of times an e-book can be borrowed and the duration of library access, while also banning other problematic terms such as prohibitions on interlibrary loans (Section 108).<sup>243</sup> Critically, this law does not mandate that publishers sell e-books to libraries or force them into any particular business relationship. It operates purely as a contract regulation measure that renders certain restrictive terms unenforceable under state consumer protection and contract law principles.<sup>244</sup> This regulatory approach represents a shift from prior failed attempts in other states which suffered from preemption problems, as the newly-enacted legislation instead focuses on ensuring that digital licensing agreements provide libraries with at least some equivalent rights as with traditional print acquisitions, without interfering with publishers’ rights to choose their customers or set prices.<sup>245</sup> Notably, the bill does not mandate that the terms of library e-book agreements preserve Fair Use under Section 107, perhaps protecting it from preemption concerns but also rendering its

---

<sup>241</sup> S.B. 1234, 2025 Leg., Reg. Sess. (Conn. 2025).

<sup>242</sup> *Id.*

<sup>243</sup> *Id.*

<sup>244</sup> Ginny Monk, *CT House Approves Bill Placing Limits on Library E-book Contracts*, CT MIRROR (May 15, 2025), <https://ctmirror.org/2025/05/15/ct-library-e-book-contracts-passes/> (Senate Bill 1234 “wouldn’t allow simultaneous restrictions on the loan period and the number of times people can borrow the book.” The bill “would also require that contracts allow inter-library loans and ban secrecy clauses that prevent libraries from talking to each other about the terms of the agreement.” The legislation operates through contract regulation, by “ban[ning] public libraries from entering into contracts with e-book publishers that contain terms that some lawmakers called restrictive” rather than mandating publisher behavior); *Libraries Call on Connecticut to Address eBook Licensing*, HARTFORD COURANT (Dec. 9, 2023), <https://www.courant.com/2023/12/09/tens-of-thousands-down-the-drain-ct-officials-aim-for-first-in-the-nation-action-to-address-ebook-costs/> (quoting Kyle Courtney, “Connecticut has a robust consumer protection statute. It has state contract law. It has contract preemption under the uniform commercial code, which allows us the ability to regulate state contracts and would equally allow us to regulate eBook contracts in the same way.”); Luke Feeney, *CT Librarians Push Bill That Would Expand E-book, Audiobook Access*, CT MIRROR (Feb. 21, 2024), <https://ctmirror.org/2024/02/21/ct-library-e-books-audiobooks-bill/>.

<sup>245</sup> *Id.*

utility more narrow.<sup>246</sup> But its impact, at least with respect to preserving Section 108 lending rights, is promising.

The Association of Research Libraries has suggested that State legislatures could go even farther, and instead enact bills specifying that *any* contract term conflicting with copyright exceptions and limitations is unenforceable under that state's contract law.<sup>247</sup> They observe, "this approach likely would not be preempted by the Copyright Act because it is not inconsistent with the exclusive rights provided to copyright owners under the Copyright Act; after all, the Copyright Act includes those exceptions."<sup>248</sup> Previous state legislation attempting to secure improved licensing terms and pricing for e-books had been litigated and found to be preempted by U.S. Copyright Law because the laws regulated rights under copyright by dictating to whom publishers must exercise distribution rights, rather than merely addressing business terms governed solely by contract.<sup>249</sup>

Arguably state procurement laws or regulations also could be proposed requiring that public universities and research institutions sign licensing agreements only if they uphold Fair Use, interlibrary loan, and TDM (or TDM plus AI) rights. Such measures should be drafted to fall within states' ability to regulate how public institutions disburse funds, akin to the state of Connecticut being able to exercise procurement authority in regulating contractual terms for public-funded libraries.<sup>250</sup> Generally speaking, states have broad constitutional authority over their own expenditures and can attach

---

<sup>246</sup> S.B. 1234, 2025 Leg., Reg. Sess. (Conn. 2025).

<sup>247</sup> Klosek, *supra* note 108.

<sup>248</sup> *Id.* at 13.

<sup>249</sup> Mike Masnick, *Unfortunate, But Not Surprising: Court Blocks Maryland's Library eBook Law*, TECHDIRT (Mar. 3, 2022), <https://www.techdirt.com/2022/03/03/unfortunate-not-surprising-court-blocks-marylands-library-e-book-law/>.

<sup>250</sup> Ellen Paul, *Libraries Need Fair eBook Contract Terms. Here's Why.*, CT MIRROR (Apr. 3, 2023), <https://ctmirror.org/2023/04/03/libraries-need-fair-ebook-contract-terms/> ("Normally, when the state or a municipality buys anything from pencils to asphalt, they go through a procurement process to ensure that taxpayer funds are being spent responsibly. Somehow, publishers have been able to skirt procurement law for the last 20 years. Librarians in Connecticut are asking for that to change... Think about California's gas mileage requirements for cars. California isn't telling Ford that they MUST sell cars to California. They're saying that if Ford WANTS to sell cars to California, they need to adhere to certain terms and conditions California sets. That is what Connecticut's eBook bills are attempting to do: regulate an out-of-control library eBook marketplace.")

conditions to the use of state funds.<sup>251</sup> When states act as market participants (buyers of goods and services) rather than as regulators, they have greater freedom to impose conditions on their contracting partners.<sup>252</sup> However, even if such bills were enacted, they likely would not reach the contracting practices of privately-funded institutions in a given state that do not come under the state's spending authority.<sup>253</sup>

In the context of consumer materials, state legislation may also be effective. One might assume Amazon and their like-minded digital content sellers have dominant market and lobbying power and will block any legislation aimed at curtailing their power over consumers at the state level. But, in 2024, California successfully passed a law mandating that sellers must not use “buy” or “purchase” to describe a transaction

<sup>251</sup> For example, in *California Teachers Assn. v. Hayes*, 5 Cal.App.4th 1513, 1531 (1992), the court explained that “the entire lawmaking authority of the state, except the people’s right of initiative and referendum, is vested in the Legislature, and that body may exercise any and all legislative powers which are not expressly or by necessary implication denied to it by the Constitution.” This plenary authority extends specifically to education funding, where the court noted that “the Legislature’s control over the public education system is still plenary” and includes the power to appropriate funds and to establish spending priorities. *Id.* at 1533, 1527. *See also* 13 Cal. Jur. 3d Constitutional Law § 125 (Westlaw current through Apr. 2025) (“It is also within the legislature’s power to impose reasonable rules and regulations governing the exercise of a constitutional power. All intendments favor the exercise of the legislature’s plenary authority. Accordingly, any doubts regarding the legislature’s power are resolved in favor of the exercise of that power or the validity of the legislature’s action.” (citing numerous cases))

<sup>252</sup> States may set specific terms, conditions, and requirements in their marketplace contracts, often tailored to their policy goals. Consider state contracting law statutes in Oregon, New York, and California as examples: ORS § 279A.015 (“a sound and responsive public contracting system should... (4) Clearly identify rules and policies that implement each of the legislatively mandated socioeconomic programs that overlay public contracting and accompany the expenditure of public funds”); ORS § 353.130 (Oregon Health and Science University shall develop contracting policies “designed to encourage affirmative action, recycling, inclusion of art in public buildings, the purchase of services and goods from individuals with disabilities, the protection of workers through the payment of prevailing wages as determined by the Bureau of Labor and Industries, the provision of workers’ compensation insurance to workers on contracts and the participation of minority-owned businesses, woman-owned businesses, veteran-owned businesses and emerging small businesses”); ORS § 279C.335 (provides criteria for granting public contracts and exemptions to public contracting requirements, emphasizing factors such as public safety, market conditions, and the complexity of public improvements); ORS § 279A.157 (allows state contracting agencies to deviate from standard contract terms if approved by the Attorney General or other relevant authorities—ensuring legal sufficiency for unique procurement needs); Cal. Pub Contract Code § 6108 (prohibits state contracts from involving the use of sweatshop labor, forced labor, convict labor, indentured labor, abusive child labor, or exploitation of children in the production or laundering of apparel, equipment, materials, or supplies); NY CLS Pub A § 2879 (public authorities and public benefit corporations may not contract with foreign business enterprises located in discriminatory jurisdictions).

<sup>253</sup> *See, e.g.*, Cal. Ed. Code § 94154 (“state will not limit, alter, or restrict the rights of private colleges, nonprofit entities, and university applicants to maintain, construct, reconstruct, and operate projects or to establish and collect rents, fees, and other charges necessary to meet operational expenses and fulfill agreements with bondholders or contracting parties”).

when digital ownership or unrestricted access is not guaranteed.<sup>254</sup> State level consumer protection laws could be introduced in other states as well to prohibit deceptive commercial practices where sellers market content as if they are similar to and substitutive of traditional physical media. Because such protections aim to more clearly describe to consumers the reality of the underlying transaction, this type of legislation unfortunately would not grant consumers new rights to digital content, but would at least raise consumers' awareness of their current dire lack of rights to digital content ostensibly in their collections.

### *B. Private Ordering*

In this section we discuss rights preservation successes resulting from libraries and the publishers working together on improved contractual terms, under a framework we refer to as “private ordering.” “Private ordering” may be thought of as market participants' voluntary efforts to address a problem, reflecting the notion that parties may be empowered to develop their own solutions to override issues through mechanisms like local policy-setting, negotiation, industry standards, or market shifts.<sup>255</sup> These private options have varying promise depending on the nature of the resource identified above—with different prospects for library-licensed materials than for consumer contracts for the reasons discussed *infra*.

#### 1. Expanding library licensing expertise and education

At least in the context of negotiated library e-resource agreements—as to which copyright owners and intermediaries are at least typically willing to negotiate (which is not the case for most consumer agreements or ToS—two factors may support better outcomes for libraries: a modest influx of lawyers working within libraries<sup>256</sup> and a general expansion in training resources available to support non-legal experts. For example, relying on licensing expertise within its library system, the University of California was able to negotiate suitable alternative terms with Springer-Nature that restored Fair Use rights for university scholars. The parties reached agreement with the following language<sup>257</sup>:

Except as explicitly stated in this Agreement or otherwise permitted in writing by Licensor as an amendment to this License Agreement, or as permitted by any Creative Commons licenses or public domain dedications applied to the Content, Customer and its Authorized Users may not:

<sup>254</sup> Cal. A.B. 2426, 2023-24 Reg. Sess. (2024).

<sup>255</sup> Pamela Samuelson, Copyright and Freedom of Expression in Historical Perspective, 10 J. INTELL. PROP. L. 319 (2003); Pam Samuelson, *Copyright, Commodification, and Censorship: Past as Prologue—But to What Future?* (presentation at Commodification of Information Conference 1999).

<sup>256</sup> See *The Journal of Copyright in Education and Librarianship*, Vol. 6, No. 1 (2022) (special issue on pathways to copyright librarianship for lawyers).

<sup>257</sup> License Agreement Between Springer-Nature and the Regents of the University of California, <https://ucop.app.box.com/s/v1260wmpe8pmm1jepwj36t7u1k1cssi8>.



(i) directly or indirectly develop, train, program, improve, and/or enrich any artificial intelligence tool (“AI Tool”), whether developed internally or provided by a third party, *unless*: (a) doing so with reasonable information security standards to undertake, mount, load, or integrate the Content on Customer’s or Authorized Users’ servers or equipment, and (b) no Content or any part thereof is shared with anyone other than Customer and its Authorized Users; and

(ii) in the case of third-party generative artificial intelligence tools, train or fine-tune any such tool (including an algorithm) *unless*: (a) doing so locally in a self-hosted or closed hosted environment solely for use by Customer or Authorized Users, and (b) none of the third-party generative artificial intelligence tool, its data, or the Content is shared with anyone other than Customer or its Authorized Users.

In no case may Customer or its Authorized Users use the Content in connection with an AI Tool, including to train or fine-tune any such tool, to: create a competing or commercial product or service; disrupt the functionality of the Content or the Platform; or reproduce or redistribute the Content or any part thereof to a third party.

Subject to the restrictions set forth in this Agreement and all applicable laws, *use of the Content in connection with AI Tools by Customer and Authorized Users is permitted (emphasis added)*.

Now all of the following AI uses are permitted under various parameters:

Table 1: Summary of Authorized AI Tools and Uses Under Negotiated License Language

Type of AI Tool	Authorized Uses
Home-grown <i>non-generative</i> ; Home-grown <i>generative</i> ; Third-party <i>Non-generative AI</i>	<b>Can be used <i>and</i> trained, provided:</b> <ul style="list-style-type: none"> <li>• Licensed content available only to licensee &amp; authorized users</li> <li>• Doesn't disrupt functioning of licensed products</li> <li>• Doesn't reproduce/redistribute licensed products to third parties</li> <li>• Doesn't create a competing or commercial product or service</li> <li>• Reasonable information security standards</li> </ul> <b>Subject to the above, <i>no</i> restrictions on dissemination of the AI tool (trained or not)</b>
Third-party <i>Generative AI</i>	<b>Can be used <i>and</i> trained provided:</b> <ul style="list-style-type: none"> <li>• Licensed content available only to licensee &amp; authorized users</li> <li>• Done only in a closed-hosted environment</li> <li>• Doesn't disrupt functioning of licensed products</li> <li>• Doesn't reproduce/redistribute licensed products to third parties</li> <li>• Doesn't create a competing or commercial product or service</li> <li>• ++Reasonable info security standards (but this would already be covered by closed-hosted environment)</li> <li>• No release or exchange of the trained tool or its data with a third party</li> </ul>

More broadly, the extensibility of such negotiating successes to other research institutions or consortia depends on expanding capacity and expertise for licensing within libraries. The license agreements that research libraries sign are highly complex and impenetrable, and many libraries lack in-house legal experts to address these complexities.<sup>258</sup> Failing to notice even just minor words in the agreement can have monumental effects on Fair Use. Consider, for example, the difference between a clause that begins, "Licensee and Authorized Users may use the Content to..." versus "Licensee and Authorized Users may **not** use the Content except to..." While legal experts in libraries have written guidance to help other institutions issue-spot and strategize,<sup>259</sup> libraries' mileage in achieving preferred contractual outcomes still varies widely.

<sup>258</sup> See *supra* note 62.

<sup>259</sup> See *supra* note 62.

One key educational component of expanding library licensing expertise means improving antitrust education within libraries and university general counsel offices. That is because one barrier to negotiating better terms may be (in our view) the risk-averse nature of university general counsel's offices in cautioning research libraries against sharing information. Libraries would benefit from discussing and developing best practices for licensing language and strategies.<sup>260</sup> However, campus general counsel's offices may feel that such discussions across institutions could violate antitrust law, even though sharing general language to preserve substantive rights does not equate with anti-competitive conduct. Counsel's admonitions have a particularly chilling effect on information sharing amongst libraries, who may also otherwise lack in-house antitrust expertise to make informed decisions. Both research libraries and campus general counsel offices would benefit from better understanding the boundaries of antitrust issues in the context of content license agreements, to lower communicative barriers.<sup>261</sup>

Finally, one reason the University of California Libraries were able to achieve the above outcomes was because the libraries realized that they needed faculty support behind them in the negotiation process.<sup>262</sup> They engaged in outreach efforts to various faculty senate committees, and then the full academic senate, all of which resulted in a strong University of California Presidential mandate supporting library efforts to protect scholars' Fair Use rights to conduct TDM and to use and train AI.<sup>263</sup> However, by and large these policy statements might be effective only to the extent they are negotiation *mandates*—that is to say, where university faculty have authorized their research libraries to walk away from signing license agreements for which such rights cannot be secured. There are examples of consortia like FinELib, negotiating on behalf of libraries throughout Finland, that have relied on a principles mandate and have indeed walked away from contracts that failed to satisfy their negotiation principles.<sup>264</sup> On a campus level, though, libraries' hands are often tied: Faculty need to support whatever outcome

---

<sup>260</sup> ARL Board Affirms Transparency as Core Operating Principle of Licensing, ASS'N OF RSCH. LIBRARY. (July 2023),

<https://www.arl.org/news/arlb-board-affirms-transparency-as-core-operating-principle-of-licensing/>

<sup>261</sup> One component of antitrust education should include encouraging research libraries to reject publishers' efforts to include non-disclosure provisions in the content license agreements. Such clauses impede "transparency and prevent information-sharing within the library sector and with third parties, such as policy-makers and funders." COMMUNIA, *supra* note 14.

<sup>262</sup> The authors engaged in providing multiple presentations to campus and UC systemwide faculty senate library committees and interest groups. See also *Anticompetitive Practices Reduce Diversity of Knowledge*, ASS'N OF RSCH. LIBRARY. (July 9, 2022), <https://www.arl.org/blog/anticompetitive-practices-reduce-diversity-of-knowledge/> ("Some libraries involve faculty in their negotiations, which can be a smart strategy to inform publishers of scholars' needs, and to help others on campus understand the complexity of these deals and contracts.")

<sup>263</sup> UNIVERSITY OF CALIFORNIA, *President Drake and Provost Newman Affirm the University's Commitment to Protect Author, Researcher and Reader Rights*, UNIV. OF CAL., (June 19, 2024), <https://ucnet.universityofcalifornia.edu/employee-news/president-drake-and-provost-newman-affirm-the-universitys-commitment-to-protect-author-researcher-and-reader-rights/>.

<sup>264</sup> COMMUNIA, *supra* note 14; FinELIB, *supra* note 15.

results from such principled engagement, and in the majority of cases faculty may decide that they cannot forgo access to content even if it means being stuck with unfavorable contractual terms.<sup>265</sup>

## 2. *Efficient breach*

Whereas licensing expertise may work for negotiated e-resources, other private ordering solutions—like the concept of “efficient breach”—may fare better for non-negotiated materials like those available on public websites or via consumer mass-market agreements. On various grounds, some scholars have suggested that researchers could choose not to adhere to the terms of an overly-restrictive agreement, and we discuss the theory of “fair breach” based on ideological principles in the judicial solutions, *infra*. However, there is also an *economic* theory of efficiency within private contracts that could be made to support non-compliance with contractual provisions, particularly for ToS or EULAs that are non-negotiable and/or that impose usage prohibitions on publicly-available public domain materials.

The concept of efficient breach suggests that a party to a contract may be well-positioned to breach that agreement if the economic benefits of doing so outweigh any damages it may owe to the non-breaching party for non-performance.<sup>266</sup> Justice Oliver Wendell Holmes is considered to have articulated the theory in 1897, writing that, “The duty to keep a contract at common law means a prediction that you must pay damages if you do not keep it—and nothing else.”<sup>267</sup> By absolving parties from any moral bedrock compelling performance, efficient breach instead frames the choice to perform as one that may be rationally determined based on economic advantage: If the damages one must pay are less than the gains one makes failing to perform, the economic calculus weighs in favor of not performing and simply making the other party whole financially.<sup>268</sup> Critics suggest that approaching contracts without a moral “conscience” (e.g. shame, or guilt) in this manner degrades “contracts as an instrument of social cooperation and debas[es] the virtues of mutuality and reciprocity that form the essence

---

<sup>265</sup> Bowers, *supra* note 10. See also *Anticompetitive Practices Reduce Diversity of Knowledge*, ASS’N OF RESEARCH LIBRARY. (July 9, 2022), <https://www.arl.org/blog/anticompetitive-practices-reduce-diversity-of-knowledge/> (“at the end of the day, libraries must make difficult decisions about the best way to support research and education within the constraints of their institution’s budget and spending priorities”).

<sup>266</sup> See HOWARD O. HUNTER, MODERN LAW OF CONTRACTS § 1:3 (database updated Mar. 2025).

<sup>267</sup> Oliver Wendell Holmes Jr., *The Path of the Law*, 10 HARV. L. REV. 457, 462 (1897).

<sup>268</sup> *BRC Rubber & Plastics, Inc. v. Continental Carbon Company*, 981 F.3d 618, 632 (7th Cir. 2020) (compiling sources articulating the efficient breach theory); *Reid Hospital and Health Care Services, Inc. v. Conifer Revenue Cycle Solutions, LLC*, 8 F.4th 642, 655, 658 (7th Cir. 2021) (“Contract law has evolved to encourage, or at least to tolerate, deliberate breaches when the breaching party will come out ahead financially if it both breaches and pays the other party damages....An efficient breacher still needs to make the non-breaching party whole.”)

of social interaction and dependency.”<sup>269</sup> Proponents tend to view it as promoting market efficiency by empowering parties to redirect resources productively.<sup>270</sup>

In the context of researchers seeking to download or use public-domain materials, efficient breach would seem to be an apt vehicle of private ordering on several grounds: What damages has a website owner suffered and what damages are they able to prove if a scholar utilizes public domain content from the site even though the ToS forbid it? The website can claim no copyright infringement. The question is whether access controls are a sufficient contractual tool. Even if the website owner suggests it would have charged a license fee for public domain content usage rights, is the website owner likely to file suit against the individual scholar, given the seemingly low anticipated return on such litigation? Ultimately, what is the likelihood that the website owner would even discover the scholar’s unauthorized use of the content?<sup>271</sup> Moreover, for those researchers doing computational research, courts have demonstrated growing skepticism toward claims of harm arising from low-impact, noncommercial scraping or reuse. In *X Corp. v. Bright Data Ltd.*,<sup>272</sup> the court dismissed X’s claims related to the act of scraping, because X could not show the scraping by itself caused any harm. Similarly, in *X Corp. v. Center for Countering Digital Hate*,<sup>273</sup> X’s claims failed because X did not allege any damages stemming directly from CCDH’s scraping. The court concluded that “it is not plausible”—nor “foreseeable”—“that this small-scale, non-commercial scraping would” cause X the harm it alleged.”

Should libraries advise scholars looking to use materials available on publicly-accessible websites that they should consider the likelihood of harm or damages in deciding whether to breach the websites’ terms? Likely invariably, libraries will likely be reluctant to encourage scholars to engage in this kind of risk calculus, even if the library is not party to the agreement and the university is not implicated in liability (i.e. even if the ToS governs the individual academic user making use of a website, not the institution). Libraries and librarians are governed by ethical codes<sup>274</sup> and likely do not wish to be seen as a marketplace actor inducing others to violate contracts; as such, they may be hesitant even to explain the efficient breach theory due to concerns over institutional reputation. They may also be concerned for the reputations of institutional scholars, given that website owners could theoretically inflict reputational harm to scholars by publicly “outing” the unauthorized downloading. Yet, regardless of whether libraries advise scholars of reputational concerns, studies have shown that for relatively

<sup>269</sup> Dawinder S. Sidhu, *The Immorality and Inefficiency of an Efficient Breach*, 8 TRANSACTIONS: TENN. J. BUS. L. 61, 99 (2006).

<sup>270</sup> See generally *supra* note 269.

<sup>271</sup> Emily Kadens, *Cheating Pays*, 119 COLUM. L. REV. 527, 589 (“Small cheats can go undetected or unpunished; reputational information can be ambiguous or ignored.”)

<sup>272</sup> 733 F.Supp.3d 832.

<sup>273</sup> 2024 WL 1246318, at 1 (N.D. Cal. March 25, 2024).

<sup>274</sup> See, e.g., ALA Code of Ethics, AM. LIBRARY. ASS’N, <https://www.ala.org/tools/ethics>

small infractions, reputational concerns are not an effective control on behavior.<sup>275</sup> Scholars may perceive little practical threat in proceeding with downloading and use, as long as their underlying access to the content was lawful.

Efficient breach is off-the-table entirely when it comes to library-licensed resources. Negotiated agreements typically contain provisions authorizing the licensor to terminate access to the resources for the entire campus—and in some cases with no refund—if it discovers that unsanctioned use has occurred that is not remedied. For example, the Springer-Nature e-books agreement signed by the University of California provides:

If Licensee breaches its obligations under these Product Terms, **Licensor may, in its sole discretion, terminate** either (i) this Text and Data Mining License only or (ii) the License Agreement, in each case in accordance with Section 7 of the General Terms and Conditions. **In addition** to, and without prejudice to any contractual rights and/or remedies under applicable law, **Licensor retains the right to suspend access to the Content in the event that Licensor determines, in its reasonable judgment, that Licensee or its Authorized Users are in breach of any of the conditions of these Product Terms.**<sup>276</sup>

Libraries which have not successfully negotiated to disclaim liability for the acts of their users may even be on the hook for contractually-mandated damages for such behavior, particularly if they are found to have been negligent or willful in failing to advise authorized users of the terms of the agreement. For example, Moody's (which offers market and research data for license to academic institutions) requires that such institutions be "responsible and liable for the compliance of...Users...with the terms and conditions of this Agreement and any breach hereof by a User...shall within the definition of 'Client' for all relevant purposes of this Agreement."<sup>277</sup> Unlike with "free" resources on the web that are used by single individuals contrary to ToS, publishers are more likely to be incentivized to pursue litigation against universities that they perceive to have deeper pockets to cover alleged damages.

In all events, academic libraries need to secure access to content for the users at their institutions. This means maintaining strong relationships with publishers and intermediaries (vendors or aggregators), and honoring agreements in good faith—a dedication to compliance that libraries have long upheld.<sup>278</sup>

---

<sup>275</sup> *Id.* See also HOWARD O. HUNTER, MODERN LAW OF CONTRACTS, § 1:3 (database updated Mar. 2025) (illustrating that, "A consumer who orders 100 items at \$1 each may be irritated when the package contains only 98, but the consumer is unlikely to pursue a \$2 claim. If the cheater follows that pattern among thousands of transactions, most of which are one-off and not repeated, the gain to the cheater may be substantial and low risk.")

<sup>276</sup> License Agreement Between Springer-Nature and the Regents of the University of California, <https://ucop.app.box.com/s/v1260wmpe8pmm1jepwj36t7u1k1cssi8>.

<sup>277</sup> See Moody's Analytics, Terms of Agreement, [https://drive.google.com/file/d/1Gy3WiL8r8BBxu5hNVQid29V1\\_dSrYb5/view?usp=sharing](https://drive.google.com/file/d/1Gy3WiL8r8BBxu5hNVQid29V1_dSrYb5/view?usp=sharing) (Par. 1).

<sup>278</sup> Ellen Finnie Duranceau, License Compliance, 26 SERIALS REV. 53 (2000).

### C. Judicial approaches

Finally, there may be some versions of these contracts that are susceptible to legal challenges in court. Though as explained above, ongoing relationships with distributors of research library content (much of which is only available through one or a few sources) may make lawsuits an undesirable option for libraries, we believe these solutions may be particularly for consumer market materials and especially for public websites ToS, especially those that rely on legal theories that include elements related to protections for consumers.

#### 1. Challenging the validity of the contract

One of the foundational reasons for contractual override taking on such importance is because so many transactions involving digital works have been thought to be subject to enforceable contracts. As mentioned in the introduction section, much of this deference given to private contracts stems from *ProCD* and decisions following *ProCD* since the 1990s, finding consumer form contracts—presented through shrinkwrap and click-wrap agreements (mostly what we see with consumer market materials) or browse-wrap agreements (present on many publicly accessible websites)—to be valid.<sup>279</sup> Judge Easterbrook seems to have sincerely believed that upholding the validity of unseen and unnegotiated form contracts is both socially efficient and beneficial, allowing big corporations to save time from negotiating individual contracts and use the saved time for nobler pursuits.<sup>280</sup> Despite its well-meaning policy goals, *ProCD*'s assessment of contract validity was one-sided.<sup>281</sup> It did not take into consideration well-established prior case law that consistently held that shrinkwraps were not enforceable and it ignored time-honored contract formation rules as stated in the Second Restatement of Contracts.<sup>282</sup> Section 211(3) Standardized Agreements states:

Where the other party has reason to believe that the party manifesting such assent would not do so if he knew that the writing contained a particular term, the term is not part of the agreement.<sup>283</sup>

The accompanying Official Comment f further clarifies the limit of standardized agreements:

Subsection (3) applies to standardized agreements the general principles stated in §§ 20 and 201. Although customers typically adhere to standardized agreements and are bound by them without even appearing to know the standard terms in detail, they are not bound

---

<sup>279</sup> *Supra* note 38.

<sup>280</sup> *Id.*

<sup>281</sup> Cheryl B. Preston and Eli W. McCann, *Unwrapping Shrinkwraps, Clickwraps, and Browsewraps: How the Law Went Wrong from Horse Traders to the Law of the Horse*, 26 *BYU J. PUB. L.* 1 (2012).

<sup>282</sup> It is one of the best-recognized and frequently cited legal treatises in all of American jurisprudence, also selectively cited by Judge Easterbrook in *ProCD*.

<sup>283</sup> Second Restatement of Contracts. Section 211(3)



to unknown terms which are beyond the range of reasonable expectation. [...] a party who adheres to the other party's standard terms does not assent to a term if the other party has reason to believe that the adhering party would not have accepted the agreement if he had known that the agreement contained the particular term. [...] Reason to believe may be inferred from the fact that the term is bizarre or oppressive, from the fact that it eviscerates the non-standard terms explicitly agreed to, or from the fact that it eliminates the dominant purpose of the transaction. The inference is reinforced if the adhering party never had an opportunity to read the term, or if it is illegible or otherwise hidden from view.<sup>284</sup>

Scholarly criticism of *ProCd* has been voluminous,<sup>285</sup> and courts are becoming more careful in evaluating whether there's a valid contract in the first place. Courts have backed away from the fiction of consent<sup>286</sup> promulgated by *ProCD* especially in the context of browsewrap.<sup>287</sup> Notably in Ninth Circuit, browsewraps generally cannot form enforceable agreements<sup>288</sup>, and even clickthrough agreements require the court to examine if the agreements are "reasonably conspicuous" so that users should be on notice.<sup>289</sup>

<sup>284</sup> *Id.*

<sup>285</sup> See Yochai Benkler, *Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74 N.Y.U. L. REV. 354, 429–35 (1999); David Nimmer et al., *The Metamorphosis of Contract into Exploit*, 87 CALIF. L. REV. 17 (1999); MARGARET J. RADIN, *BOILERPLATE: THE FINE PRINT, VANISHING RIGHTS, AND THE RULE OF LAW* 39–50, 168–76 (2013) (boilerplate in IP contracts degrade democracy by employing mass contracts to restructure and supersede the rights given by legislators, thereby deleting rights granted by the democratic process); *Techy Tuesday - Running Windows X?*, THE SILICON GRAYBEARD (Sept. 8, 2015), <https://thesilicongraybeard.blogspot.com/2015/09/techy-tuesday-running-windows-x.html>. But see Guy Rub, *Copyright Survives: Rethinking the Copyright-Contract Conflict*, 2017 VA. L. REV. 1141 (2017). See also *Enforceability of "Clickwrap" or "Shrinkwrap" Agreements Common in Computer Software, Hardware, and Internet Transactions*, 106 A.L.R. 5TH A.L.R.5th 309 (summarizing case law).

<sup>286</sup> Jonathan A. Obar & Anne Oeldorf-Hirsch, *The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services*, 23 INFO., COMM. & SOC'Y 1020 (2020) (in a click-through agreement: "Average adult reading speed (250–280 words per minute), suggests TOS [takes] 15–17 minutes to read. ... All participants were presented the TOS and had an average reading time of 51 seconds. ... 93% [agreed] to TOS, with decliners reading ... TOS 90 seconds longer.")

<sup>287</sup> For a more comprehensive discussion of enforceability of clickthrough, browsewrap, sign-in wrap and scrollwrap contracts, see Kieran McCarthy, *A Comprehensive Guide to the Law of Online Contracts*, MCCARTHY LAW GROUP, <https://mccarthylg.com/a-comprehensive-guide-to-the-law-of-online-contracts/>

<sup>288</sup> *Berman v. Freedom Financial Network*, 30 F.4th 849 (9th Cir. 2022); *Sellers v. JustAnswer*, 73 Cal. App. 5th 444 (2021); *Long v. Provide Commerce*, 245 Cal. App. 4th 855 (2016); and *Nguyen v. Barnes & Noble*, 763 F.3d 1171 (9th Cir. 2014).

<sup>289</sup> *Meyer v. Uber Technologies, Inc.*, 868 F.3d 66 (2d Cir. 2017) (applying 9th Cir. law; "whether the Terms of Service is 'reasonably conspicuous' requires considering many factors, the analysis focuses primarily on the design of site where the contract is formed. The three appellate judges reviewed the screen size of the plaintiff's Samsung Galaxy S5 and how the screenshot of the final step of the registration process appears on that screen, the presence of words and other things in

Broadly speaking, browsewrap agreements—where terms are posted on a website but no affirmative action is required to accept them—are presumptively unenforceable unless the user has actual or constructive notice of the terms. Courts have regularly scrutinized these agreements due to their lack of visible or affirmative consent. In *Specht v. Netscape Communications Corp.*, the Second Circuit refused to enforce a browsewrap arbitration clause because the user would not have seen the terms without scrolling past a download button.<sup>290</sup> Similarly, in *Nguyen v. Barnes & Noble, Inc.*, the Ninth Circuit held that merely placing a hyperlink at the bottom of a webpage was insufficient to put users on notice.<sup>291</sup> These cases show that for a browsewrap to be valid, courts require clear evidence that users were aware of the terms and continued use after receiving notice.

Clickwrap agreements, where users affirmatively click “I agree” to accept terms, are generally enforceable, but can be challenged when the process fails to provide adequate notice or is otherwise procedurally unfair. Courts have upheld clickwraps in cases like *Davidson & Assocs. v. Jung*, which emphasize the significance of an affirmative act of assent.<sup>292</sup> However, not all clickwraps are automatically valid. In *Grosvenor v. Qwest Comm’ns Int’l, Inc.*, the court refused to enforce a clickwrap because the arbitration clause was buried behind multiple layers of navigation and was not clearly incorporated into the agreement.<sup>293</sup> This case shows that a clickwrap may be invalid if essential terms are not reasonably accessible at the time of acceptance or if the user’s assent is ambiguous due to interface design or contradictory language.

Examining the specific terms embodied in an online contract can also lead to other grounds for a finding that no valid contract exists. In *Meta Platforms, Inc. v. Bright Data Ltd.*, the court found that Meta’s ToS only applied to registered users, not all visitors.<sup>294</sup> This conclusion was reached because “The 2009 Facebook Terms included the following clause: ‘accessing or using our website . . . signif[ies] that you . . . agree to be bound by these Terms . . . , whether or not you are a registered member of Facebook.’ . . . Therefore, it is reasonable to infer that the current Terms contemplate a “user” as an account holder.” So when Bright Data scraped without logging in, it was not a user bound by Meta’s ToS. This reasoning still applied for the time period when Bright Data maintained accounts on Meta platforms, because the accounts were not used as a part of the web scraping (they were only used for marketing).

Researchers and archivists routinely access, scrape, use, and preserve publicly available materials in service of scholarship, cultural memory, and the public interest. As courts continue to scrutinize the enforceability of online contracts, the implications for these activities are substantial. Many of the most valuable sources for public interest research are only accessible via public websites or the consumer market. These

---

close proximity to the “agree” button, and the appearance, location, and interpretive function of the agree button itself.”)

<sup>290</sup> 306 F.3d 17 (2d Cir. 2002).

<sup>291</sup> 763 F.3d 1171 (9th Cir. 2014).

<sup>292</sup> 422 F.3d 630 (8th Cir. 2005).

<sup>293</sup> 2010 WL 3906253 (D. Colo. Sept. 30, 2010).

<sup>294</sup> 2024 WL 251406 (N.D. Cal. Jan. 23, 2024). Meta never appealed this ruling and settled the case shortly after. Interestingly, Meta previously hired Bright Data to scrape other people’s websites, the exact same behavior it complained of in this case.

platforms and vendors often deploy browsewrap or clickwrap agreements to assert contractual control that extends well beyond the limited rights granted to rightsholders under copyright law. When the terms prohibit activities that would otherwise be lawful, the case law suggests that one powerful defense could be that the contracts are not valid in the first place. Crucially, when researchers and libraries reuse content without creating accounts or seeing and affirmatively agreeing to these restrictive terms, they are in all likelihood not bound by the restrictive contractual terms. When researchers can show that the terms were not reasonably conspicuous or they had no meaningful notice or opportunity to consent, courts are not likely to hold the researchers liable for breach of contract.

## 2. Copyright preemption challenges

Another promising approach may be to challenge certain contracts on the grounds that they are preempted by federal copyright law. Preemption stems from the Constitution's Supremacy Clause, which states that “the Laws of the United States . . . shall be the supreme law of the land.”<sup>295</sup> State laws that conflict with federal law are preempted. Contracts of course are private agreements, not state laws themselves, but when the power of the state through its court is used to enforce those contracts, courts have held that state action occurs and is subject to possible preemption.<sup>296</sup>

Not every contract that affects use of copyrighted works will be preempted. In fact, in many cases, courts have declined to find that state contract claims are preempted by federal copyright law, but there are a few exceptions and some recent indication that courts may be willing to take a broader look at preemption particularly in the context of non-negotiable form contracts and perhaps some negotiated consumer media contracts as well.<sup>297</sup>

It's helpful to first start with the two pathways to preemption that the Supreme Court has identified: (1) Express preemption, where Congress just says that a certain type of state law is preempted, or (2) implied preemption.<sup>298</sup> Implied preemption can occur in two instances: where Congress has chosen to “occupy the field” (field preemption) or where there is “any conflict with a federal statute” (conflict preemption).<sup>299</sup> Conflict preemption can further be divided into situations “where compliance with both state and federal regulations is an impossibility” or “those instances where the challenged state law ‘stands as an obstacle to the accomplishment

<sup>295</sup> U.S. CONST., art. VI, cl. 2. See also *Hillsborough Cnty., Fla. v. Automated Med. Lab'ys., Inc.*, 471 U.S. 707, 712–13 (1985) (“It is a familiar and well-established principle that the Supremacy Clause, U.S. CONST., art. VI, cl. 2, invalidates state laws that “interfere with, or are contrary to,” federal law.”) (quoting *Gibbons v. Ogden*, 22 U.S. 1, 211 (1824)).

<sup>296</sup> *Shelley v. Kraemer*, 334 U.S. 1 (1948) (holding that state enforcement of private agreements—in this case, racially restrictive covenants—are subject to the Federal law).

<sup>297</sup> Guy A. Rub, *Moving from Express Preemption to Conflict Preemption in Scrutinizing Contracts over Copyrighted Goods*, 56 AKRON L. REV. 56, 303 (2023).

<sup>298</sup> *Arizona v. United States*, 567 US 387, 399 (2012).

<sup>299</sup> See *Oneok, Inc. v. Learjet, Inc.* 575 U.S. 373 (2015).

and execution of the full purposes and objectives of Congress.”<sup>300</sup> Field preemption can also be understood as a form of conflict preemption.<sup>301</sup>

a. Express preemption challenges

Congress has chosen to exercise its power to expressly preempt state and common laws that provide for “legal or equitable rights that are equivalent to any of the exclusive rights within the general scope of copyright.”<sup>302</sup> The scope of copyright is defined by Sections 102-122 of Copyright Act, which identifies copyrightable subject matter, exclusive rights of copyright owners, and limitations on those rights such as Fair Use. It seems like a simple proposal, that if something falls within the scope of copyright, and the rights granted under state law deal with the “equivalent” of such copyright right, then the state law claim is preempted. Yet, since copyright preemption was first codified, courts have struggled to determine whether the rights created by contract are “*equivalent to any of the exclusive rights within the general scope of copyright.*”

Like its holdings regarding contract validity, *ProCD* was also influential in shaping the copyright preemption jurisprudence.<sup>303</sup> There, Judge Easterbrook reasoned that “[a] copyright is a right against the world. Contracts, by contrast, generally affect only their parties; strangers may do as they please, so contracts do not create ‘exclusive rights.’”<sup>304</sup> Similarly, several other circuits have also concluded that rights provided by contract are not generally “equivalent” to those under the copyright act, generally because they contain a minimal “extra element”—even just an exchange of promises.<sup>305</sup>

By contrast, other Circuits have taken a more permissive approach to preemption. Most notably, the Second Circuit has developed jurisprudence that strongly favors preemption. For example, the Second Circuit in *Harper & Row Publishers, Inc. v. Nation Enterprises*,<sup>306</sup> the court stated that “if unauthorized publication is the gravamen of [the plaintiffs’] claim, then it is clear that the right they seek to protect is coextensive

<sup>300</sup> *Arizona v. United States*, 567 U.S. 387, 399 (2012).

<sup>301</sup> *English v. Gen. Elec. Co.*, 496 U.S. 72, 79 (1990). (“field pre-emption may be understood as a species of conflict preemption,” because “[a] state law that falls within a pre-empted field conflicts with Congress’ intent . . . to exclude state regulation.”)

<sup>302</sup> 17 U.S.C. § 301(a).

<sup>303</sup> Viva R. Mofatt, *Super-Copyright: Contracts, Preemption, and the Structure of Copyright Policymaking*, 41 U.C. DAVIS L. REV. 45 (2007); Caleb Nelson, *Preemption*, 86 VA. L. REV. 225, 226–29 (2000) (the Supreme Court’s “[preemption] taxonomy recognizes three different types of preemption: ‘express’ preemption, (implied) ‘field’ preemption, and ‘conflict’ preemption”).

<sup>304</sup> *ProCD, Inc. v. Zeidenberg*, 86 F.3d 1447, 1454 (7th Cir. 1996). But it is increasingly the case that what is supposed to be “private” contracts end up controlling broad public behavior, especially with technological advancement, anti-circumvention legislation, and growing monopolies in the market, see Mark A. Lemley, *Beyond Preemption: the Law and Policy of Intellectual Property*, 87 CAL. L. REV. 111 (1999) (“technology [allows] a vendor to interpose contract terms even in a downstream transaction that would not ordinarily be thought to demonstrate privity between the ‘contracting’ parties.”).

<sup>305</sup> *Nat’l. Car Rental Sys., Inc. v. Comp. Assoc. Int’l., Inc.*, 991 F.2d 426, 433 (8th Cir. 1993); *Taquino v. Teledyne Monarch Rubber*, 893 F.2d 1488, 1501 (5th Cir. 1990); *Acorn Structures, Inc. v. Swantz*, 846 F.2d 923, 926 (4th Cir. 1988).

<sup>306</sup> 723 F.2d 195, 201 (2d Cir. 1983) (reversed on other grounds, 471 U.S. 539 (1985).)

with an exclusive right already safeguarded by the Act—namely, control over reproduction and derivative use of copyrighted material.” In the Second Circuit where there exists a strong tradition of copyright preemption, courts are dismissing claims based on copyright preemption even in recent AI cases.<sup>307</sup>

Other circuits have rejected *ProCD*’s categorical approach, favoring a case-by-case analysis that sometimes results in courts finding a variety of state laws, including application of state contract law, preempted. For example, in *Lasercomb America Inc v Reynolds*, the Fourth Circuit refused to uphold the following contract term: “Licensee agrees during the term of this Agreement [99 years] and for one (1) year after the termination of this Agreement, that it will not write, develop, produce or sell or assist others in the writing, developing, producing or selling computer assisted die making software, directly or indirectly without Lasercomb’s prior written consent.”<sup>308</sup> Even post-*ProCD*, the Sixth Circuit in *Wrench Ltd. Liab. Co. v. Taco Bell Corp.*,<sup>309</sup> held that contracts must have some meaningful additional elements: “if the [contractual] promise amounts only to a promise to refrain from reproducing, performing, distributing or displaying the work, then the contract claim is preempted.” By contrast, courts in circuits still heavily influenced by *ProCD* may not easily dismiss claims based on preemption.<sup>310</sup>

Some contracts may well be subject to a preemption challenge just because they do not actually provide such an extra element. For example, the Second Circuit in *ML Genius v. Google* was asked to weigh in on the validity of a contract claim ML Genius made against Google for violating its website ToS, which prohibited users from using the Genius site “to copy, modify, sell and/or distribute content appearing on Genius’s website.”<sup>311</sup> ML Genius operates a website where users upload transcriptions of song lyrics, and the website makes money by selling ads on its website. A third party, LyricFind, was copying those transcriptions and licensing them to Google so Google could display them in user search results. This meant that ML Genius was deprived of web traffic and therefore ad revenue.

The Second Circuit in evaluating the claim reminded us that the “extra element” analytical framework is workable so long as one doesn’t apply it mechanically:

The critical inquiry is whether such extra elements of the state law claim beyond what is required for copyright infringement change the nature of the action so that it is *qualitatively* different from a copyright infringement claim. To determine whether a claim is qualitatively different, we look at what the plaintiff seeks to protect, the theories in which the matter is thought to be protected and the rights sought to be enforced. The “extra element” inquiry is not “mechanical.” While we have inquired into the existence

<sup>307</sup> *New York Times Co. v. Microsoft Corp.*, 2025 WL 1009179 (S.D.N.Y. Apr. 4, 2025) (granting motion to dismiss).

<sup>308</sup> 911 F.2d 970 (4th Cir. 1990).

<sup>309</sup> 256 F.3d 446 (6th Cir. 2001).

<sup>310</sup> See Kieran McCarthy, *Should Copyright Preemption Moot Anti-Scraping TOS Terms?*, TECH. AND MARKETING L. BLOG (Dec. 13, 2023), <https://blog.ericgoldman.org/archives/2023/12/should-copyright-preemption-moot-anti-scraping-tos-terms-guest-blog-post.htm>.

<sup>311</sup> 2022 WL 710744 (2d Cir. Mar. 10, 2022).

of extra elements in determining whether preemption applies, this inquiry “requires a holistic evaluation of the nature of the rights sought to be enforced, and a determination whether the state law action is qualitatively different from a copyright infringement claim.”<sup>312</sup>

In the context of contracts that limit a user’s exercise of Fair Use or other copyright users rights, courts have been asked to consider whether those limits amount to a state law intrusion on “equivalent rights” under the Copyright Act. For example, *Bowers v. Baystate Technologies*, in which the Federal Circuit was asked to pass on the validity of a shrink-wrap license that prohibited all reverse engineering of Bowers’ software.<sup>313</sup> Baystate, which was being sued for breach of that agreement, asserted that such contractual restrictions were invalid because they were preempted by the Copyright Act, and more specifically because they purported to restrict an activity that the Fair Use defense would have otherwise supported.<sup>314</sup> The court was not persuaded. Starting from the proposition that “courts respect freedom of contract and do not lightly set aside freely-entered agreements,” and that in general, the courts have allowed “contractual waiver of affirmative defenses and rights,” the Bowers court recounted the by-then familiar logic that contracts are distinct from copyright’s exclusive rights because they require the extra element of promise and consideration. It found nothing exceptional about the fact that the terms of this contract happened to limit uses otherwise acceptable under Fair Use, and held the contract enforceable.

The majority opinion in *Bowers* was not without dissent. Judge Dyk, concurring in part and dissenting in part, articulated the same basic policy concerns we highlight throughout this article: “The majority’s approach permits state law to eviscerate an important federal copyright policy reflected in the fair use defense, and the majority’s logic threatens other federal copyright policies as well.”<sup>315</sup> While Judge Dyk agreed with the majority that a state can permit parties to contract away a Fair Use defense if the contract is freely negotiated, he reasoned that because states are not permitted to eliminate the Fair Use defense, neither can they authorize copyright owners to eliminate the defense themselves through mass application of shrinkwrap licenses that are offered on essentially “take it or leave it” ToS. While Judge Dyk’s policy objections have received some attention, several other circuits have followed the approach adopted by *Bowers*.<sup>316</sup>

---

<sup>312</sup> *Id.*

<sup>313</sup> 320 F.3d 1317, 1325 (Fed. Cir. 2003). This case was brought in the District Court of Massachusetts and the contract was governed by Massachusetts law. The Federal Circuit heard this copyright case because it was associated with a patent infringement claim. In deciding the copyright preemption issue, the Federal Circuit was obliged to decide based on what it understood First Circuit precedent to mean.

<sup>314</sup> *Id.* (citing *Atari Games Corp. v. Nintendo of America, Inc.*, 975 F.2d 832 (Fed. Cir. 1992) (finding fair use for reverse engineering)).

<sup>315</sup> *Id.* at 1335. (J. Dyk, concurring in part and dissenting in part).

<sup>316</sup> See also *Davidson & Assoc. v. Jung*, 422 F.3d 630, 639 (8th Cir. 2005) (“While *Bowers* and *Nat’l Car Rental* were express preemption cases rather than conflict preemption, their reasoning applies here with equal force. By signing the TOUs and EULAs, Appellants expressly relinquished their rights to reverse engineer.”). See also *White v. Samsung Electronics America, Inc.*, 989 F. 2d 1512 (9th Cir. 1993) (J. Kozinski, dissenting from the order rejecting the

That's not to say that *Bowers* is the only approach to express preemption in cases involving contractual override. One of the earliest cases on point, a Fifth Circuit decision in *Vault Corp. v. Quaid Software*,<sup>317</sup> actually supports the idea that contract terms which limit user protections should not be enforced because they are expressly preempted by federal law. In *Vault*, the plaintiffs sold computer diskettes that contained software designed to prevent unauthorized duplication of their contents. Use of the diskettes was governed by a license agreement that prohibited a wide variety of copying. The contract was governed by Louisiana law, which includes a specific statute that authorized software contract terms to prohibit copying for any purpose.<sup>318</sup> Quaid devised a way to defeat Vault's protections, but in doing so made a number of copies of Vault's software by decompiling and disassembling Vault's program. Among other things, Vault sued Quaid for breaching the license agreement. Quaid's defense was that the license restrictions are invalid because the Copyright Act explicitly permits such activities under Section 117— a specific exception allowing users to make adaptations as long as the copying is an essential step in utilizing a computer program. Quoting the Supreme Court for the point that "[w]hen state law touches upon the area of [patent or copyright statutes], it is 'familiar doctrine' that the federal policy 'may not be set at naught, or its benefits denied' by the state law,"<sup>319</sup> the Fifth Circuit held that because the Louisiana law permits a software producer to limit rights of computer program owners under Section 117, that aspect of state law was preempted and therefore the license restriction is unenforceable.

#### b. Conflict preemption challenges

Looking beyond express preemption, two recent cases make us think another approach to preemption, conflict preemption, may hold promise—particularly for contract terms of mass-applicability, with no opportunity to negotiate, that attempt to exert exclusive control over information primarily for economic reasons.

First, the Second Circuit's decision in *In re Jackson*,<sup>320</sup> provides a compelling framework for applying implied conflict preemption when state action overrides Fair Use and other exceptions. Adjudicating a claim arising under Connecticut's right of publicity law, the court addressed a dispute between rappers 50 Cent (aka Curtis James Jackson, III) and Rick Ross (aka William Leonard Roberts II) related to Ross' sampling from 50 Cent's song "In Da Club" and his use of 50 Cent's name in a remix album. 50 Cent alleged that this violated his right of publicity under Connecticut common law. Judge Leval (writing for the court), affirmed the dismissal on the grounds of conflict preemption. The *Jackson* court relied on underlying standards set forth in *Bonito Boats v. Thundercraft Boats*<sup>321</sup>, for the principle that, "When a person undertakes to exert

---

suggestion for rehearing en banc) (objecting to the majority's refusal to rehear a preemption challenge to California's right of publicity statute and explaining why the conflict with Fair Use should result in the statute being preempted).

<sup>317</sup> 847 F.2d 255 (5th Cir. 1988).

<sup>318</sup> La. Rev. Stat. Ann. § 51:1961 *et seq.* (1987).

<sup>319</sup> 847 F.2d at 269 (quoting *Sears, Roebuck & Co. v. Stiffel Co.*, 376 U.S. 225, 229 (1964)).

<sup>320</sup> 972 F.3d 25, 26 (2020).

<sup>321</sup> 489 U.S. 141 (1989)



control over a work within the subject matter of the Copyright Act under a state law mechanism that is different from the one instituted by the law of copyright, implied preemption may bar the claim unless the state-created right vindicates a substantial state law interest”—specifically, an interest “outside the sphere of Congressional concern in the [copyright] laws.”<sup>322</sup>

To that end, the court established the rule that courts are more likely to find preemption “the less substantial the state law rights invoked” or “the more the invocation of the state right amounts to little more than camouflage for an attempt to exercise control over the exploitation of a copyright.”<sup>323</sup> 50 Cent’s claim was preempted via both prongs: The state lacked a substantial state interest in enforcing the statute because the right of publicity claim was not vindicating any rights as distinct from those furthered by federal copyright law, and further the claim was a “little more than a thinly disguised effort to exert control over an unauthorized production of a sample of [50 Cent’s] work.”

*In re Jackson* can be applied to private contracts that abrogate Fair Use or other copyright exceptions. Conflict preemption prevents states from using their judicial machinery—whether through statutes or contract enforcement—to enable private parties to circumvent the system Congress created in the absence of substantial contrary interests.<sup>324</sup> States lack substantial interests in enforcing contractual provisions that completely eliminate Fair Use. States certainly have legitimate interests in protecting private investment—such as for publishers who digitize public domain materials—and in promoting financial remuneration for labor or costs invested in publishing. However, when private contracts eliminate Fair Use of copyrighted works or foreclose the use of public domain materials entirely, rather than simply requiring compensation for publishers or licensors, they exceed legitimate protections for investment and instead recreate the exclusive property right or control that copyright law specifically limits. The Constitution’s mandate for the progress of science and useful arts, and Congress’ mandates for a robust public domain and preservation of user rights, are foundational to the balance struck in the copyright ecosystem—and outweigh state commercial interests, particularly when compensation could protect investment without eliminating Federally-guaranteed usage rights and limitations on copyright monopolies.

This brings us to the second important case, from the Northern District of California, *X Corp. v. Bright Data*<sup>325</sup>, which is the first to actually apply the framework outlined in *Jackson* to a contract claim. The dispute in *X Corp* is straightforward: X Corp, a social media company, brought a breach of contract claim (and several other claims) against a webscraping company for scraping and then reselling publicly available data—primarily posts created by users on the platform—in violation of its ToS.<sup>326</sup> There, the court concluded that “the extent to which public data may be freely copied from social media platforms, even under the banner of scraping, should generally

---

<sup>322</sup> *Id.* at 37 (citing *Bonito Boats*, 489 U.S. at 155).

<sup>323</sup> *Id.*

<sup>324</sup> *Shelley v. Kraemer*, 334 U.S. 1 (1948).

<sup>325</sup> 733 F. Supp. 3d 832.

<sup>326</sup> *Id.* at 832.

be governed by the Copyright Act, not by conflicting, ubiquitous terms.”<sup>327</sup> The X Corp court identified three reasons why it came to this conclusion.

First, it observed that as a non-exclusive licensee of user generated content, X Corp is not actually a copyright owner or an exclusive licensee, and therefore is not permitted under the Copyright Act to exclude others from reproduction or distribution of those works.<sup>328</sup> Yet, its ToS purport to give it precisely those rights by virtue of state contract law.

Second, the court directly addressed how X Corp’s ToS limit users access to Fair Use: “Although Section 106 gives a copyright owner exclusive rights to do and to authorize the reproduction, adaptation, distribution, and display of their copyrighted works, Section 107 provides for the exception that *anyone* may make fair use of copyrighted works without permission or payment of money. 17 U.S.C. § 107. This statutory privilege may or may not apply in any given instance, but in *all* instances it would be obliterated by X Corp.”<sup>329</sup> Such an exercise, the court concluded, would “flout[] Congress’s intent that “[t]he *limited* scope of the *copyright holder’s* statutory monopoly ... reflect[] a balance of competing claims upon the public interest.”<sup>330</sup>

Third, observing that Congress has placed limits on the scope of copyright protection: X Corp. would give itself *de facto* copyright ownership over content that Congress declined to extend copyright protection to in the first place (*e.g.*, likes, user names, short comments) when that content, “not ‘original’ in the constitutional sense[,] ... may not be copyrighted.” *Feist Publications, Inc. v. Rural Tel. Serv. Co.*. This shrinks the public domain, restricting free reproduction, adaptation, distribution, and display of publicly available, non-expressive material.”<sup>331</sup>

The impact of these two cases shouldn’t be overstated: the framework in *In Re Jackson* has been applied only a handful of times, and *X Corp* is a lone district court decision without broad precedential effect. And yet, these cases do make for a promising and persuasive argument that in at least some circumstances—especially with website ToS covering user-generated content, or even with non-negotiable consumer market material EULAs that include blanket restrictions on downstream use that would “obliterate” even the possibility of a fair use assertion—courts have good reason to be skeptical of enforcing those terms. In practical terms,

### 3. Other judicial strategies and defenses

Finally, there are a handful of other potential strategies that litigants might use to challenge contracts that may be applicable in some special circumstances. Some other context-specific defenses that have worked based on the behavior of the copyright owner.. For example, for webscraping, *hiQ Labs, Inc. v. LinkedIn Corp.*,<sup>332</sup> highlighted a

<sup>327</sup> *Id.* at 851.

<sup>328</sup> *Id.* at 848-49.

<sup>329</sup> *Id.* at 852.

<sup>330</sup> *Id.* (quoting *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 431 (1984) (emphasis added)).

<sup>331</sup> *Id.* (quoting *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 347-48 (1991)).

<sup>332</sup> 639 F.Supp.3d 944 (N.D. Cal., 2022).

number of potential defenses. There the court accepted HiQ's "waiver" defense, stating that a reasonable jury could find waiver a valid defense if "aside from generally deploying anti-scraping technology against all scrapers, [LinkedIn] did not take steps to legally enforce against known scraping by hiQ for years, and it allowed its employees to attend hiQ's conferences."<sup>333</sup> Somewhat similar to the waiver defense, the *hiQ* court also viewed favorably an estoppel argument where the scraper can show that they relied on a public website's knowing acquiescence as it developed its scraping business. There, the court observed that LinkedIn employees attended a *hiQ* conference did not timely object to its business model—this could be enough to justify hiQ's "near total dependency on LinkedIn."

These defenses are highly dependent on the behavior of the copyright or website owner and its relationship with the user, yet they are common enough that they are worth noting here given the commonalities with both library and research behavior, which often includes open and collaborative use of web content, often with knowledge of the copyright owner.

There may be other viable strategies that are not as well developed. For example, Professor Margaret Chon has written about the application of copyright misuse and why it may have particular salience for libraries. The misuse doctrine, traditionally applied to patent law, suggests that copyright owners who use their rights to engage in anticompetitive behavior or to stifle Fair Use by extending their rights through improper licensing.<sup>334</sup> Chon gives an extensive review of the ways in which modern media licensing terms have hindered the traditional public interest role of libraries and suggests that "contractual terms that eliminate exhaustion, raise prices beyond fair remuneration, and/or eliminate consumer freedoms such as fair use."<sup>335</sup>

Another approach is to develop an equitable "fair breach" defense, a right conceived of as analogous to Fair Use. Professor Jane Ginsburg, speculating in the 1990s that "when 'we're all connected,' no functional difference may exist between a contract and a property right" and that such a state may make it "necessary to consider whether limitations incorporated in the copyright law should be imported to its contractual substitute. With respect to libraries and their users, one should inquire whether some kind of fair use exception is appropriate. This might take the form of a judge-made right of 'fair breach.'"<sup>336</sup> Other scholars have advocated for similar approaches, such as Professor Julie Cohen, in the context of proposals to extend the UCC to provide licensors with special rights as against purchasers.<sup>337</sup> Professor Mark Lemley likewise observes that courts have already limited the construction of contracts

---

<sup>333</sup> *Id.*

<sup>334</sup> See Mark Lemley, *Beyond Preemption: The Federal Law and Policy of Intellectual Property Licensing*, 87 CAL. L. REV. 111, 151 (1999) (discussing copyright misuse as one of three additional limitations on licensing, beyond copyright preemption).

<sup>335</sup> Margaret Chon, *Protecting Progress: Copyright's Common Law and Libraries* 72 J. Copyright Society 761 (2025).

<sup>336</sup> Ginsburg, *supra* note 29.

<sup>337</sup> Julie E. Cohen, *Copyright and the Jurisprudence of Self-Help*, 13 BERKELEY TECH. L. J. 1089 (1998).

under state law when they might otherwise conflict with federal policy,<sup>338</sup> pointing for example to cases such as *S.O.S., Inc. v. Payday, Inc.*<sup>339</sup> where the court held that the license at issue “must be construed in accordance with the purposes underlying federal copyright law,” and that therefore the court would “rely on state law to provide the canons of contractual construction, but only to the extent such rules do not interfere with federal copyright law or policy.”<sup>340</sup> Lemley thus argues that courts might well build on those precedents to hold a “‘no-reverse-engineering,’ ‘no criticism,’ or ‘no-parody’ clause in a contract [ ] as unenforceable as a matter of federal policy, without invoking the mechanisms of preemption.”<sup>341</sup>

The precise contours of “fair breach” defense would need to be defined by courts, though it seems obvious that the areas of most serious public policy conflict would likely arise in cases of contracts of adhesion, applied at a mass scale, and particularly when the seller has such market power to effectively replace the default rules of copyright with its own. Professor Pamela Samuelson and Kurt Opsahl, writing in the context of the then-proposed Article 2B of the UCC governing information transactions, identified three particular concerns—contracts that restrict innovation, that limit free expression, and that inhibit competition—that could help guide such an inquiry.<sup>342</sup>

The possibility of judicial intervention is probably best suited for challenges to consumer market EULAs and website ToS. Both present stronger cases when considering all of the possible defenses outlined here. Both also avoid many of the business issues that would complicate a case under research library contracts (for instance, that universities are repeat players and must retain a relationship with vendors who in some cases hold a virtual monopoly over content in certain publishing areas).

One of the challenges with judicial resolution is identifying the appropriate party to bring a lawsuit: who will sue? Because the companies in control of EULAs and ToS are also typically in control of when and against whom to initiate legal action, they can structure claims in a way that places consumer-defendants at a disadvantage. Big companies are often likely reluctant to enforce contracts in a way that would undermine the perceived legitimacy or enforceability of their contracts. The aggressive litigation strategy pioneered by the music industry in the 2000s has proven ineffective in stopping unwanted behaviors such as piracy, and at the same time brought negative publicity for the big companies.<sup>343</sup>

<sup>338</sup> Mark A. Lemley, *Beyond Preemption: the Law and Policy of Intellectual Property*, 87 CAL. L. REV. 11, 161 (1999).

<sup>339</sup> 886 F.2d 1081, 1088 (9th Cir. 1989).

<sup>340</sup> *Id.* at 1088. See also *SQL Solutions, Inc. v. Oracle Corp.*, 1991 U.S. Dist. LEXIS 21097, at \*7 (N.D. Cal. Dec. 18, 1991) (“[F]ederalism principles dictate that state rules of contractual construction cannot interfere with federal law or policy.”).

<sup>341</sup> Lemley, *supra* note 201 at 163.

<sup>342</sup> Pamela Samuelson & Kurt Opsahl, *Licensing Information in the Global Information Market: Freedom of Contract Meets Public Policy*, 21 EUR. INTELL. PROP. REV. 386 (1999).

<sup>343</sup> See, e.g., Eric Bangeman, *Judge’s decision leaves RIAA with lose-lose situation in Elektra v. Santangelo*, ARSTECHNICA, (March 21, 2007), <https://arstechnica.com/tech-policy/2007/03/judges-decision-leaves-riaa-with-lose-lose-situation-in-elektra-v-santangelo/>.

All of the judicial pathways identified above are ultimately defenses that leave researchers and libraries in an uncomfortable and uncertain position.<sup>344</sup> Users may be able to file declaratory judgements, where they may choose the forum and framing of the dispute. The consumers would be able to preemptively seek a court's ruling on the legality or enforceability of a contract. To demonstrate Article III standing, a party must demonstrate it has "(1) suffered an injury in fact, (2) that is fairly traceable to the challenged conduct of the defendant, and (3) that is likely to be redressed by a favorable judicial decision."<sup>345</sup> Although declaratory judgment actions may be possible for some plaintiffs, they still require a "real and substantial" dispute, not merely hypothetical or speculative concerns. Courts have consistently required that the harm be imminent and concrete in order to satisfy the requirement under Article III.<sup>346</sup> Only rarely can a party have standing to vindicate a nonparty's rights.<sup>347</sup> Even when not successful on the merits, declaratory actions can raise awareness of the contract override problem and contribute to shaping the future narrative around consumer rights.

Judicial solutions may be ineffective in most cases given imbalance of power of the parties involved. Because of this imbalance of power, one plausible option may be to emphasize to state attorneys general the challenges and ask state attorneys general to pursue an investigation. Many of the potential defenses we discussed above are rooted essentially in consumer protection concerns—namely that large companies are exploiting their market power in a way that undermines established legal rights and public policy objectives. By highlighting the potential for widespread consumer harm and the erosion of Fair Use, advocates can encourage attorneys general to scrutinize these contracts and potentially initiate legal challenges. Furthermore, focusing on the imbalance of bargaining power inherent in many consumer contracts can strengthen arguments that these agreements are unconscionable.

---

<sup>344</sup> Some courts have allowed defendants to assert copyright misuse as a counterclaim, but given the essential elements of a copyright misuse claim it seems unlikely that it could be successfully asserted as a standalone claim outside of a declaratory judgment action. *See Amaretto Ranch Breedables, LLC v. Ozimals, Inc.*, 790 F.Supp.2d 1024, 1033 (N.D. Cal. 2011) (collecting cases permitting and prohibiting copyright misuse as a counterclaim or independent claim).

<sup>345</sup> *Spokeo, Inc. v. Robins*, 578 U.S. 330, 338 (2016).

<sup>346</sup> *MedImmune, Inc. v. Genentech, Inc.*, 549 U.S. 118 (2007). Contrast the consumer's position when filing a declaratory judgment with the business's position when suing for breach of contract: the consumer will need to prove some concrete harm, but many courts would consider "breach of contract" alone to suffice as concrete damages suffered by a business, see *Springer v. Cleveland Clinic Emp. Health Plan Total Care*, 900 F.3d 284, 292 (6th Cir. 2018) (finding concrete injury in breach of contract alone because, "[l]ike any private contract claim, his [Article III] injury does not depend on allegation of financial loss. His injury is that he was denied the benefit of his bargain."), *J.P. Morgan Chase Bank, N.A. v. McDonald*, 760 F.3d 646, 650–51 (7th Cir. 2014) ("When one party fails to honor its commitments, the other party to the contract suffers a legal injury sufficient to create standing even where that party seems not to have incurred monetary loss or other concrete harm."), *Katz v. Pershing, LLC*, 672 F.3d 64, 72 (1st Cir. 2012) (damages are sufficiently shown "when a plaintiff generally alleges the existence of a contract, express or implied, and a concomitant breach of that contract").

<sup>347</sup> *McCollum v. Cal. Dep't of Corr. & Rehab.*, 647 F.3d 870, 879 (9th Cir. 2011) ("[A] plaintiff must show his own injury, a close relationship between himself and the parties whose rights he asserts, and the inability of the parties to assert their own rights.")

Case law is developing in recent years, addressing whether ToS can override the public's right to make Fair Use or otherwise lawful reuse of content hosted on public-facing websites, and it seems reasonable to anticipate continued interest to litigate in this space by plaintiffs and defendants alike<sup>348</sup>. An increase in disputes over ToS is especially likely given that the common invocation of the Computer Fraud and Abuse Act (CFAA) to ban data scraping on public sites—including rescinding permission through banning IP addresses or sending cease and desist letters—has now proven unreliable.<sup>349</sup> Public interest organizations and legal scholars should watch the space closely and be prepared to file amicus briefs in support of Fair Use and equitable access to data.

## CONCLUSION

Copyright limitations and exceptions matter for research: they result in a higher production of new works of scholarship<sup>350</sup>, and drive emerging methodologies as a means of extracting information and advancing knowledge.<sup>351</sup> Traditional and new outcomes of creation are precisely what the Framers intended in imposing the “limited times” to bring all works eventually into the public domain; and “promot[ing] the progress of science and useful arts” mandates. The Framers recognized that authors, and in our case, scholars and teachers, need to be able to utilize existing knowledge resources to create new knowledge goods. Indeed, the availability of openly accessible scholarship during the COVID-19 pandemic directly benefited public health policy- and decision-making.<sup>352</sup> Congress and the Copyright Office clearly understood the importance of facilitating research access and usage rights for science and libraries,

<sup>348</sup> Some scholars speculate that big companies may be less willing to litigate web-scraping cases, because they do not want to create anti-scraping case law that could be used against them when they are the ones undertaking the web scraping. For example, see Kieran McCarthy, *Facebook Drops Anti-Scraping Lawsuit Against Bright Data*, TECH & MARKETING L. BLOG (Feb. 28, 2024), <https://blog.ericgoldman.org/archives/2024/02/facebook-drops-anti-scraping-lawsuit-against-bright-data-guest-blog-post.htm>.

<sup>349</sup> In 2017, after receiving a cease and desist letter from LinkedIn citing potential CFAA liability, hiQ sought—and was granted—a preliminary injunction to prevent LinkedIn from blocking hiQ's access to public LinkedIn profiles. In *hiQ Labs, Inc. v. LinkedIn Corp.*, 31 F.4th 1180 (9th Cir. 2022), the court held that CFAA applied narrowly to non-public, password-protected data, not public-facing data. The 9th Circuit Court expressed concerns about large platforms monopolizing public data, which “would disserve the public interest.” However, outside of CFAA, the court said very little—it did not address whether website ToS can prevent a scraper's use of public data. However, when the case was sent back down to the lower court, it was revealed that hiQ may have accessed non-public data, making LinkedIn's CFAA claim relevant once more. By April 2021, hiQ was no longer in business. The parties settled at the end of 2022.

<sup>350</sup> Mike Palmedo, *The Impact of Copyright Exceptions for Research on Scholarly Output*, 2 EFIL J. OF ECON. RESEARCH 114, <https://ssrn.com/abstract=3090022> (2019).

<sup>351</sup> Christin Handke, Lucie Guibault, & Joan-Josep Vallbé, *Copyright's impact on data mining in academic research*, 42 MANAGERIAL AND DECISION ECON. 1999, <https://doi.org/10.1002/mde.3354> (2021).

<sup>352</sup> Yian Yin et al., *Coevolution of Policy and Science During the Pandemic*, 371 SCIENCE 128 (2021).

having implemented Sections 107 and 108 without any statutory or regulatory exclusions or opt-outs. Yet, some copyright owners are using contracts as an end-run around these essential research exceptions.

Addressing this issue in the context of its impact on libraries and research is timely: More and more content necessary for research is being siphoned off exclusively into licenses, or is subject to contracts, rather than being made available through traditional purchases that allow for perpetual access, archiving, interlibrary loan, and ownership by libraries.<sup>353</sup> In many cases, there are no other access options other than to consent to the publisher or licensor's restrictive terms.<sup>354</sup> Though research and library exceptions function as a means to balance copyright owners' interests with societal opportunities for scientific advances, they cannot keep up: The pervasiveness of mass-market contracts, and the market power differential between publishers and libraries, have now squarely overtaken scholars' and libraries' ability to rely on the fundamental rights that Congress afforded.<sup>355</sup>

A sweeping EU-style Federal statute that protects research- and library-related limitations and exceptions from contractual override would provide the most comprehensive solution, regardless of the nature of the content agreement. Achieving such a reform will be extremely challenging, but it remains possible—especially if the proposed exception is sympathetic and narrowly tailored (e.g., right to repair statutes), and if stakeholders remain optimistic and work proactively with key government agencies such as the U.S. Copyright Office.

All other solutions discussed in this paper may hold more limited value. Their efficacy depends instead on the type of contract at issue and whether it can be negotiated. For example, in the context of non-negotiated ToS and EULAs, state legislation along the lines of the Connecticut e-book bill may be a viable, though time-consuming, advocacy pathway for libraries seeking to address non-negotiable agreements. Similarly, the *X Corp. v. Bright Data*<sup>356</sup> case offers some hope that state attorneys general may pursue judicial remedies against contracts of adhesion that undermine public policy. However, the court system likely will continue to dissuade engagement from resource-poor libraries and scholars seeking individual relief from oppressive e-resource agreements. For individual researchers relying on Fair Use, certain private ordering mechanisms do exist: They may consider efficient breach of non-negotiable agreements for online content that they have lawfully accessed—especially where the contractual override neither serves legitimate interests nor results in any actual harm. Yet, this is a solution challenging for institutions to contemplate (much less to encourage) from a policy and risk perspective.

For library license agreements, and absent the desired comprehensive Federal legislation foreclosing override of limitations and exceptions in research contexts, the most viable pathway forward rests on expanding legal expertise within libraries to understand, negotiate, and advocate against problematic provisions. Despite budget constraints, libraries will need to be resourced to, and build capacity to, engage in

---

<sup>353</sup> Bowers, *supra* note 10.

<sup>354</sup> *Id.*; see also COMMUNIA, *supra* note 14; McCracken, *supra* note 79.

<sup>355</sup> *Id.*

<sup>356</sup> 733 F.Supp.3d 832 (N.D. Cal. 2024).



collective action, with careful understanding of the limits of antitrust constraints. With e-resource agreements, the landscape for preserving research rights in the digital age seems navigable only through this form of sustained, coordinated private action, rather than holding out for legislative or judicial salvation. Still, even large negotiating consortia have at times proven no match for the strength that copyright owners and intermediaries may wield. Private ordering has demonstrable benefits for library-negotiated contracts, but is not a plenary solution that would resolve concerns in the manner that an improved-upon European Union framework would.