AI and Doctrinal Collapse
78 STAN. L. REV. (forthcoming 2026)

ARTICLE

AI AND DOCTRINAL COLLAPSE

ALICIA SOLOW-NIEDERMAN†

Abstract. Artificial intelligence runs on data. But the two legal regimes that govern data—information privacy law and copyright law—are under pressure. Formally, each regime demands different things. Functionally, the boundaries between them are blurring, and their distinct rules and logics are becoming illegible.

This Article identifies this phenomenon, which I call "inter-regime doctrinal collapse," and exposes the individual and institutional consequences. Through analysis of pending litigation, discovery disputes, and licensing agreements, this Article highlights two dominant exploitation tactics enabled by collapse: Companies "buy" data through business-to-business deals that sidestep individual privacy interests, or "ask" users for broad consent through privacy policies and terms of service that leverage notice-and-choice frameworks. Left unchecked, the data acquisition status quo favors established corporate players and impedes law's ability to constrain the arbitrary exercise of private power.

Doctrinal collapse poses a fundamental challenge to the rule of law. When a leading AI developer can simultaneously argue that data is public enough to scrape—diffusing privacy and copyright controversies—and private enough to keep secret—avoiding disclosure or oversight of its training data—something has gone seriously awry with how law constrains power. To manage these costs and preserve space for salutary innovation, we need a law of collapse. This Article offers institutional responses, drawn from conflict of laws and legal pluralism, to create one.

† Associate Professor of Law, The George Washington University. Thank you to BJ Ard, Elettra Bietti, Bob Brauneis, Danielle Keats Citron, Julie Cohen, David Freeman Engstrom, Kat Geddes, Jake Goldenfein, Aziz Huq, Amanda Levendowski, Martha Minow, Chris Morten, Paul Ohm, Richard Re, Barak Richman, Jon Siegel, Jessica Silbey, Ben Sobel, Dan Solove, Xiyin Tang, Salomé Viljoen, Ari Ezra Waldman, Daniel Wilf-Townsend, and participants at the 75th Annual ICA Conference Panel, Code is Not Law: Reassembling the Social in Technology Law and the 2025 Privacy Law Scholars Conference and attendees of the George Washington University Law School faculty workshop for invaluable comments, conversations, and support. Thank you to Rachel Layne and Rhyia Bibby for outstanding research assistance. I also wish to thank the editors of the Stanford Law Review, especially Boyce Buchanan and Emily Harrington, for comments and questions that helped me to refine and sharpen the argument. This paper was awarded an honorable mention in the 2026 AALS Scholarly Papers Competition. The Article was substantively finalized in December 2025, and reflects updates through that point in time. All remaining errors and omissions are my own.

Table of Contents

Introduction

Artificial intelligence (AI) is stressing the law—but not in the way that you might think. Despite myriad warnings that data-driven technologies are outpacing legal regulation,[1] the fast clip of technological development is not the most important issue.[2] The fundamental source of strain isn't timing or technology per se; rather, it's the law. More precisely, AI is stressing the law because of longstanding choices about how to structure the fields of law that regulate data, and the ways that these choices enable AI companies today to leverage longstanding doctrinal tensions.

This Article identifies an underlying source of legal strain, which I call "inter-regime doctrinal collapse" ("doctrinal collapse" or "collapse" for short),[3] and uses the example of AI and data acquisition to illustrate collapse on the ground. Because AI models rely on data, and because data is governed by two legal domains, information privacy law and intellectual property law (primarily copyright law), there is overlapping coverage of the same regulatory object. If the privacy-copyright boundary does not remain sufficiently distinct, and the discrete rules and logics of each domain are not legible, then the two regimes lose their independent structural integrity and collapse into one another.[4]

Whether inter-regime doctrinal collapse is good or bad depends on what sorts of behavior it enables. Collapse sounds catastrophic, but it is not always a negative phenomenon. Sometimes, legal structures are weak, outdated, inadequate, or otherwise ripe for reconstruction; moreover, sophisticated actors may blur doctrinal lines in creative, innovative, or otherwise beneficial ways.[5]

I contend that collapse becomes problematic when it disproportionately facilitates exploitation by already-established corporate players and impedes

---

[1] *See, e.g.*, Gary E. Marchant, *The Growing Gap Between Emerging Technologies and the Law*, *in* THE GROWING GAP BETWEEN EMERGING TECHNOLOGIES AND LEGAL-ETHICAL OVERSIGHT: THE PACING PROBLEM 19, 19 (Gary E. Marchant, Braden R. Allenby & Joseph R Herkert, eds., 2011); *Is the Law Playing Catch-Up with AI?*, HARV. L. TODAY (Jan. 16, 2025), https://hls.harvard.edu/today/is-the-law-playing-catch-up-with-ai/.

[2] *See* Margot E. Kaminski & Meg Leta Jones, *Constructing AI Speech*, 133 YALE L. J. F. 1212, 1213 (2024) (rejecting the pacing problem critique and eschewing technological determinism).

[3] As discussed *infra* Part I, others have previously developed accounts of *intra*-regime doctrinal collapse and considered how the prospect of copyright liability shapes the behavior of companies seeking to develop AI systems. To my knowledge, I am the first to theorize "doctrinal collapse" in terms that focus on *inter*-regime legal blurring and instability, with an eye to consequences for the broader legal system.

[4] This Article focuses on copyright law as an especially pressing example of this form of inter-regime doctrinal collapse. *See* discussion *infra* note 70 and accompanying text.

[5] See discussion *infra* Part I.A.

law's ability to constrain the arbitrary exercise of private power.[6] That's precisely what is happening in AI. For example, the leading generative AI company OpenAI has argued that the data it used to train its models are "public" and thus not subject to either copyright or privacy restrictions, yet the company also refuses to disclose the same material on the grounds that the data are proprietary and confidential.[7] This sort of doctrinal switching—between copyright and information privacy logics—can be understood as a form of corporate opportunism. And it is.[8] But that is not all that it is. Analyzing only the result (opportunistic behavior) overlooks the relationship between legal regimes (the inter-regime doctrinal collapse) that enables that result. Doctrinal collapse is a structural condition that warrants distinct recognition.[9]

This Article situates AI governance challenges in these terms and argues that doctrinal collapse is different from regulatory confusion, regulatory arbitrage, or strategic gap-seeking.[10] These phenomena are second-order effects that involve the behavior of private actors. Doctrinal collapse highlights the underlying conditions that facilitate that behavior: It is about the values and doctrines within a particular field of law, and how that field relates to another part of the legal system when both apply to the same regulated object. Collapse can remain a dormant underlying state until powerful private actors leverage the conditions that it creates, which in turn intensifies the collapse of the two fields into one another. This is not a speculative possibility. It is already happening with information privacy law and copyright law, and it is already playing out in lawsuits across the nation and in regulatory battles across the globe.[11] The time to act is now.

Regulatory attention is vital because, when it comes to AI, the exploitation enabled by doctrinal collapse has a theoretical and practical cost. Without sufficiently clear boundaries between different fields of law, it is not possible to detect which issues are properly controlled by which legal domain, nor to discern how particular normative rationales apply. Picture, for example, a newspaper (or an author, or another content creator) suing a generative AI company in a copyright infringement lawsuit. How do copyright law and privacy law apply to data that subsequent users of the AI tool produce as they interact with the company's service? The plaintiff might argue that is essential to preserve all content and all "output logs" generated by users, on the grounds that this

[6] *See* discussion *infra* Parts III.B.

[7] See discussion *infra* Part II.B.

[8] See discussion *infra* Part II and Part III.B.

[9] See discussion *infra* Part I.

[10] *See id.*

[11] See discussion *infra* Parts I–II.

information is relevant for copyright law claims.[12] The defendant AI company might respond by emphasizing that the demand violates user privacy and contravenes privacy laws around the world.[13] Notably, the AI company might invoke user privacy interests in this lawsuit, even when its own scraping of the internet to obtain data to train its models violates third-party privacy interests, and even if it separately invokes copyright law's fair use doctrine as a defense in generative AI litigation.[14]

How to straighten out this tangled mess of copyright and privacy interests? There's no easy answer. And that's the point. Where there is doctrinal collapse, there is a risk that "issue spotting" becomes an exercise in creative legal argumentation, rather than principled application of law. Some amount of legal ambiguity, overlap, and flexibility is inevitable, and can even be desirable, and I do not claim that law must have total clarity to remain principled.[15] But there are limits: When doctrinal lines collapse into one another, law can become unpredictable, inconsistent, and manipulable.[16] And when there is too much ambiguity, sophisticated private actors can leverage inter-regime doctrinal collapse in ways that compromise law's public legibility and legitimacy and threaten the rule of law itself.[17]

Now, these complex dynamics are not entirely new. Legal scholars have long recognized that privacy and copyright have a complicated relationship.[18]

---

[12] *See* New York Times Co. v. Microsoft Corp. et al., 1:23-cv-11195 (S.D.N.Y. Jan. 13, 2025), Doc. No. 379, at 1 ("News Plaintiffs made it abundantly clear to OpenAI that output logs are relevant to the claims . . . The Times's initial Complaint repeatedly asserts that the output of OpenAI's products based on the GPT large language models directly infringes The Times's copyrights . . . ." (internal citations and emphasis omitted)). *See also* Brad Lightcap, *How We're Responding to the New York Times' Data Demands in Order to Protect User Privacy*, OPENAI (June 5, 2025), https://openai.com/index/response-to-nyt-data-demands/.

[13] *See In re* OpenAI, Inc., Copyright Infringement Litig., 1:23-cv-11195 (S.D.N.Y. May 13, 2023), Doc. No. 551, at 1 (quoting OpenAI's statements during a court conference: "OpenAI expressed a reluctance for a 'carte blanche, preserve everything request,' . . . and raised not only user preferences and requests, but also 'numerous privacy laws and regulations throughout the country and the world . . ." (internal citations omitted)). See also discussion *infra* Part III.B.

[14] See discussion *infra* Part II.

[15] See discussion *infra* Part I.

[16] See discussion *infra* Parts II.B–C and Part III.A.

[17] See discussion *infra* Part III.B.

[18] The literature here is voluminous. *See, e.g.*, JESSICA SILBEY, AGAINST PROGRESS: INTELLECTUAL PROPERTY AND FUNDAMENTAL VALUES IN THE INTERNET AGE 156–213 (2022); Madelyn Rose Sanfilippo, Brett M. Frischmann, & Katherine J. Strandburg, *Privacy and Knowledge Commons*, *in* GOVERNING PRIVACY IN KNOWLEDGE COMMONS 5–50 (Madelyn Rose Sanfilippo, Brett M. Frischmann, & Katherine J. Strandburg, eds. 2021); DANIEL J. SOLOVE, THE FUTURE OF REPUTATION: GOSSIP, RUMOR, AND PRIVACY ON THE INTERNET 184–86 (2007) [hereinafter SOLOVE, THE FUTURE OF REPUTATION]; Sharon K. Sandeen, *Relative Privacy: What Privacy Advocates Can Learn from Trade Secret Law*, 2006 MICH. STATE L.

Well before the advent of generative AI and chatbots, a voluminous literature debated the proper way to understand the interaction between information privacy law and intellectual property law.[19] In addition, scholars have analyzed the relationship between copyright and First Amendment law[20] as well as copyright and antitrust law[21] and examined how generative AI puts pressure on these legal regimes.[22] Scholars have also focused on the ways in which copyright law mediates access to data for AI developers. For instance, Amanda Levendowski argues that "copyright law causes friction that limits access to training data and restricts who can use certain data," thereby acting as "a

---

REV. 667, 670 (2006); Julie E. Cohen, *DRM and Privacy*, 18 BERKELEY TECH. L.J. 575 (2003); Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373 (2000); Mark A. Lemley, *Private Property: A Comment on Professor Samuelson's Contribution*, 52 STAN. L. REV. 1545 (2000); Jessica Litman, *Information Privacy/Information Property*, 52 STAN. L. REV. 1283 (2000); Pamela Samuelson, *Privacy as Intellectual Property?*, 52 STAN. L. REV. 1125 (2000); Jonathan Zittrain, *What the Publisher Can Teach the Patient: Intellectual Property and Privacy in an Era of Trusted Privication*, 52 STAN. L. REV. 1201 (2000); Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105 (1990); Jon O. Newman, *Copyright Law and the Protection of Privacy*, 12 COLUMBIA J.L. & ARTS 459 (1988). *See also* Derek E. Bambauer, *Exposed*, 98 MINN. L. REV. 2025 (2014) (focusing on relationship between copyright law and privacy law in the context of non-consensual distribution of intimate imagery); Amanda Levendowski, *Using Copyright to Combat Revenge Porn*, 3 N.Y.U. J. INTELL. PROP. & ENT. L. 422 (2014) (same); Rebecca Tushnet, *How Many Wrongs Make a Copyright?*, 98 MINN. L. REV. 2346 (2014) (same).

[19] *See* Amanda Levendowski Tepski, *Fairer Public Benefit in Copyright Law*, 47 CARDOZO L. REV. 119, 129–30 n.42 (2025) (compiling sources on both sides of the debate). For recent works arguing for linkages of the two domains, see generally Cathay Y.N. Smith, *Weaponizing Copyright*, 35 HARV. J.L. & TECH. 193 (2021); Shyamkrishna Balganesh, *Privative Copyright*, 73 VAND. L. REV. 1 (2020); Andrew Gilden, *Sex, Death, and Intellectual Property*, 32 HARV. J.L. & TECH. 67 (2018); Margaret Chon, *Copyright's Other Functions*, 15 CHI.-KENT J. INTELL. PROP. 364 (2016). For recent works arguing against linkages of the two domains, see generally Benjamin L.W. Sobel, *A New Common Law of Web Scraping*, 25 LEWIS & CLARK L. REV. 147 (2021); Eric Goldman & Jessica Silbey, *Copyright's Memory Hole*, 2019 BYU L. REV. 929; Christopher Buccafusco & David Fagundes, *The Moral Psychology of Copyright Infringement*, 100 MINN. L. REV. 2433 (2016); M. Margaret McKeown, *Censorship in the Guise of Authorship: Harmonizing Copyright and the First Amendment*, 15 CHI.-KENT J. INTELL. PROP. 1 (2016); Jeanne C. Fromer, *Should the Law Care Why Intellectual Property Rights Have Been Asserted?*, 53 HOUS. L. REV. 549 (2015); *see also* Edward Lee, *Suspect Assertions of Copyright*, 15 CHI.-KENT J. INTELL. PROP. 379, 381–82 (2016) (arguing that "copyright law legitimately protects an author's reputation or privacy interests" in cases where "the author of the work is asserting the copyright").

[20] *See* Margot E. Kaminski, *Authorship, Disrupted: AI Authors in Copyright and First Amendment Law*, 51 U.C. DAVIS L. REV. 589, 596, 606 (2017) (analyzing different treatment of algorithmic authorship in copyright law and First Amendment law).

[21] *See* Jacob Noti-Victor & Xiyin Tang, *Antitrust Regulation of Copyright Markets*, 101 WASH. U. L. REV. 851, 856–57 (2024) (arguing that "antitrust and copyright law cannot work in silos" and proposing a regulatory model to address copyright market concentration).

[22] *See, e.g.*, Daryl Lim & Peter K. Yu, *The Antitrust-Copyright Interface in the Age of Generative Artificial Intelligence*, 74 EMORY L. J. 847, 849 (2025) (citing Noti-Victor & Tang, *supra* note 21, at 858)).

significant contributor to biased AI."[23] Furthermore, scholars have recently analyzed generative AI and the status of training data as a political and economic object. For example, Jake Goldenfein contends that contestation over "datasets" and their status as "content" or "data" permits competing claims about "the AI dataset as a legal and economic object" in the digital economy.[24] Furthermore, a rapidly growing literature, much of it building from Julie Cohen's deep body of work on "informational capitalism,"[25] critically assesses and theorizes platform firms' moves to generate value from information as it flows between producers, online platforms, advertisers, users, and other consumers.[26] Still missing, however, is a comprehensive framework for understanding these issues as *inter*-regime conflicts and sites of contestation, and for recognizing why the challenge of governing across multiple legal regimes matters for both individuals and for the rule of law itself.

This Article provides that cross-cutting analysis and offers a way forward. Descriptively, I offer the conceptual vocabulary—inter-regime doctrinal collapse—to better explain interactions across the fields of law that govern data. By focusing on AI and data acquisition, I expose concrete, contemporary examples of what is happening, as well as highlight the specific tactics that leading firms are using to leverage the inter-regime doctrinal collapse of privacy law and copyright law.

Normatively, I argue that the AI development status quo both prioritizes economic incumbents and makes it harder to govern private conduct in principled and predictable ways.[27] Left unmanaged, inter-regime doctrinal collapse in AI makes the path of the law unpredictable and uniquely susceptible to manipulation by sophisticated actors. This exploitation, enabled by doctrinal collapse, threatens law's capacity to constrain private power and weakens law's claim to public legitimacy. Although this rule of law problem is not fundamentally new or unique to data acquisition, contemporary AI development is a force multiplier of doctrinal collapse's causes and consequences. If we wish to regulate AI and sustain law's basic role in democratic governance, then we must create legal systems that can respond to collapse and ensure that law governs emerging, data-driven technologies in the public interest.

---

[23] Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579, 589 (2018).

[24] Jake Goldenfein, *Data or Content? The Conceptual Battles Defining Dataset Markets*, 2 PLATFORMS & SOC. 1, 2 (2025).

[25] *See, e.g.*, JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019) [hereinafter COHEN, BETWEEN TRUTH AND POWER].

[26] Among many, many possible works, see, for example, Elettra Bietti, *Data is Infrastructure*, 26 THEORETICAL INQUIRIES L. 55, 56, 58 (2025); Amanda Parsons & Salomé Viljoen, *Valuing Social Data*, 124 COLUM. L. REV. 993, 996 (2024); Salomé Viljoen, *A Relational Theory of Data Governance,* 131 YALE L.J. 573, 577 (2021).

[27] See discussion *infra* Part III.

This Article proceeds in four parts. Part II develops the concept of doctrinal collapse, distinguishing it from other kinds of doctrinal instability, like overlap or drift, and from other dynamics, like regulatory arbitrage. Part III applies that theory to what is happening in the world. It focuses on AI development and the data acquisition that supports it, illustrating how the boundaries of information privacy and copyright law are blurring and how the structural integrity of the regimes is collapsing. Part IV evaluates the broader consequences of doctrinal collapse, arguing that the exploitation it enables in AI development has considerable costs for individuals and for legal legitimacy. Part V contends that collapse is not a problem to solve, but rather a dynamic to recognize and manage. It proposes two institutional pathways—drawing from conflict of laws and legal pluralism—to preserve law's coherence and legibility and bolster rule of law values, without sacrificing regulatory flexibility and adaptability.

Although this Article focuses on AI development, the theoretical framework it develops has broader implications. Inter-regime doctrinal collapse captures how legal domains become unstable when multiple fields attempt to govern the same regulatory object. This dynamic may arise elsewhere, whether due to the contested status of those objects (as with data[28]) or due to structural tensions between overlapping regimes (as in antitrust and consumer protection, or the interaction between Section 230 and the First Amendment). By offering a grounded account of collapse in the AI and data acquisition context, this Article provides a missing vocabulary to identify and craft responses to similar breakdowns across other areas of law.

I. Defining Doctrinal Collapse

This Part first defines inter-regime doctrinal collapse and then clarifies how it differs from other regulatory phenomena, such as regulatory arbitrage and regulatory gaps. This exposition builds a theoretical foundation for Part II's analysis of doctrinal collapse in the AI development context and Part III's assessment of the political economy and rule of law stakes.

A. Doctrinal Collapse, Inside and Out

Effective legal governance of the data-driven information society demands a fresh understanding of doctrinal collapse that accounts for interactions *across* legal fields. This understanding is complementary to, but distinct from, a rich body of work analyzing doctrinal collapse *within* various fields of law. From tort

[28] Thank you to Elettra Bietti for helpful comments on this point.

law;[29] to IP law;[30] to the First Amendment;[31] to the Fourth Amendment;[32] to other areas of constitutional law,[33] scholars have assessed how a singular legal domain can erode, unravel, or otherwise become a shell of its former self.[34] There are many possible drivers of such *intra*-domain collapse. For instance, judges may conflate or merge elements of a cause of action;[35] combine previously independent lines of a doctrine;[36] or apply the doctrine in ways that expose its functional inadequacy.[37] Each of these forms of collapse occurs as the judiciary confronts a particular field of law and responds in ways that lead that singular legal domain to evolve and shift, thereby undermining a prior doctrinal pattern, unsettling an asserted doctrinal justification, or both.

A separate body of prior work considers a different form of doctrinal evolution and examines "drift" in the legal system. Much of the work on "drift" focuses on a singular area of law and emphasizes the role of the court. There are many drivers of such drift. For instance, over time, courts' treatment of particular

---

[29] *See* Nicolas P. Terry, *Collapsing Torts*, 25 CONN. L. REV. 717, 717–18 (1993); James A. Henderson, Jr. & Aaron D. Twerski, *Doctrinal Collapse in Products Liability: The Empty Shell of Failure to Warn*, 65 N.Y.U. L. REV. 265, 326 (1990).

[30] For a recent article that uses the term "boundary collapse" to analyze "doctrinal borrowing" and "boundary collapse" between patent law and copyright law, see Mark Bartholomew & John Tehranian, *Historical Kinship & Categorical Mischief: The Use and Misuse of Doctrinal Borrowing in Intellectual Property Law*, 109 IOWA L. REV. 51, 69, 84–93 (2023). For a recent account that analyzes the blurring of copyright and trademark law in the generative AI context, without using the specific term "collapse," see Katrina Geddes, *The New Art Forgers*, 58 ARIZ. ST. L. J. (forthcoming 2026), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5561522 (manuscript at 58-76).

[31] *See* Christopher S. Yoo, *The Rise and Demise of the Technology-Specific Approach to the First Amendment*, 91 GEO. L.J. 245, 283-84, 355 (2003); Susan H. Williams, *Content Discrimination and the First Amendment*, 139 U. PENN. L. REV. 615, 619 (1991).

[32] *See* Joshua A. Engel, *Doctrinal Collapse: Smart Phones Cause Courts to Reconsider Fourth Amendment Searches of Electronic Devices*, 41 U. MEMPHIS L. REV. 233, 291 (2010).

[33] *See* Ari Ezra Waldman, *Manufacturing Uncertainty in Constitutional Law*, 91 FORDHAM L. REV. 2249, 2252-54 (2023); David E. Bernstein, *The Due Process Right to Pursue a Lawful Occupation: A Brighter Future Ahead?,* 126 YALE L.J.F. 287, 289, 292–93 (2016).

[34] Moreover, First Amendment scholars have suggested, without using the term collapse, that a field of law can expand in ways that alter the doctrine and, potentially, threaten its core values. On First Amendment expansionism, see, for example, Margot E. Kaminski, *Privacy and the Right to Record*, 97 B.U. L. REV. 167, 172–74 (2017); Leslie Kendrick, *First Amendment Expansionism*, 56 WM. & MARY L. REV. 1199, 1200 (2015). On threats to core First Amendment values, see, for example, Robert C. Post, *The Classic First Amendment Tradition Under Stress: Freedom of Speech and the University*, *in* THE FREE SPEECH CENTURY 110 (Lee C. Bollinger & Geoffrey R. Stone, eds. 2019); Morgan N. Weiland, *Expanding the Periphery and Threatening the Core: The Ascendant Libertarian Speech Tradition*, 69 STAN. L. REV. 1389, 1394 (2017).

[35] *See* Terry, *supra* note 29, at 717–18, 730.

[36] *See* Williams, *supra* note 29, at 619; Bernstein, *supra* note 33, at 289.

[37] *See* Henderson, Jr. & Twerski, *supra* note 29, at 277–78.

kinds of legal claims may shift[38] or the analysis of the elements of a claim may change.[39] Additionally, courts may apply legal standards in ways that disfavor certain categories of litigants[40] or apply a doctrine in ways that compromise its original purpose;[41] alternatively, courts may respond to litigation pressures that create productive "mutation" of the original doctrine.[42] Drift can also occur across multiple levels of the judiciary; for example, decisions by a higher court may fail to provide adequate doctrinal guidance for lower courts, producing drift and instability in the lower courts.[43] Moreover, other scholarship on "drift" looks beyond the courts to foreground cultural and political factors and their impact on legal understandings, over time.[44] All of these accounts of drift reveal how legal concepts and doctrines evolve and change as they reflect, and interact with, forces both inside and outside of the legal system.

Doctrinal collapse, in the sense developed in this Article, goes beyond intra-domain accounts and builds from prior understandings of legal developments as mediated by social and political power.[45] By definition, inter-regime doctrinal collapse is an *inter*-domain phenomenon. It can occur when two legal domains each apply to a particular regulatory object (in the sense of either a literal thing—

[38] *See, e.g.*, Dmitry Karshtedt et al., *The Death of the Genus Claim*, 35 HARV. J.L. & TECH. 1, 4 (2021).

[39] *See, e.g.*, Oren Bracha, *Not De Minimis: (Improper) Appropriation in Copyright*, 68 AM. U. L. REV. 139, 143–45 (2018).

[40] *See* Katie R. Eyer, *That's Not Discrimination: American Beliefs and the Limits of Anti-Discrimination Law*, 96 MINN. L. REV. 1275, 1332 & nn.189–90 (2012).

[41] *See* Shani Shisha, *The Folklore of Copyright Procedure*, 36 HARV. J.L. & TECH. 62, 100–01, 109–10 (2022).

[42] Thomas P. Schmidt, *Standing Between Private Parties*, 2024 WIS. L. REV. 1, 17 (2024) (quoting Thomas W. Merrill, *Article III, Agency Adjudication, and the Origins of the Appellate Review Model of Administrative Law*, 111 COLUM. L. REV. 939, 972 (2011)).

[43] *See* Todd Phillips & Beau J. Baumann, *The Major Question Doctrine's Domain*, 89 BROOKLYN L. REV. 747, 753 (2024) (discussing "doctrinal drift" and asserting that "scholars have begun demonstrating how recent Roberts Court decisions are so open ended that they have created doctrinal instability in the lower courts") (citing Jacob D. Charles, *The Dead Hand of a Silent Past: Bruen, Gun Rights, and the Shackles of History*, 73 DUKE L.J. 67, 78 (2024). *But see* Richard M. Re, *Narrowing Supreme Court Precedent from Below*, 104 GEORGETOWN L. REV. 921, 925 (2016) (arguing, without discussing doctrinal drift, that "in many situations, a lower court can legitimately narrow Supreme Court precedent by adopting a reasonable reading of it").

[44] *See, e.g.*, J.M. Balkin, *Ideological Drift and the Struggle Over Meaning*, 25 CONN. L. REV. 869, 870 (1993) (describing "ideological drift:" "Styles of legal argument, theories of jurisprudence, and theories of constitutional interpretation do not have a fixed normative or political valence. Their valence varies over time as they are applied and understood repeatedly in new contexts and situations."). *See also* J.M. Balkin, *Ideological Drift*, *in* ACTION AND AGENCY 13 (Roberta Kevelson ed., 1990); J.M. Balkin, *The Promise of Legal Semiotics*, 69 TEX. L. REV. 1831, 1833 (1991); J.M. Balkin, *Some Realism about Pluralism: Legal Realist Approaches to the First Amendment*, 1990 DUKE L.J. 375, 383–85.

[45] *See supra* text accompanying notes 10–17, 44 and sources cited therein.

an object—or in the sense of realizing a goal). At least partial *overlap* of this sort is a necessary condition for inter-regime doctrinal collapse. I argue that this is the case with information privacy law and copyright law: Each regime applies to data, and each regime features distinct doctrinal and normative lines. But overlap alone is not sufficient.

Focusing on AI and data governance reveals additional requirements for doctrinal collapse: The doctrinal boundaries between the two overlapping legal domains must *blur* and the two blurred domains must feature *irreconcilable* animating logics.[46] Doctrinal boundaries can blur in multiple ways: conceptually (where categories lose meaning), functionally (where fields fail to guide or constrain conduct), and/or institutionally (where legal enforcement becomes incoherent).

Because many legal disputes about data acquisition are pending in court as of this writing,[47] leaving institutional enforcement an open question, this Article focuses on conceptual and functional blurring. I emphasize a particular model of collapse in which structural weaknesses within one domain (here, information privacy law) create conditions for the boundaries between partially-overlapping, irreconcilable regimes to blur. Specifically, privacy law's latent structural weaknesses, such as its longstanding reliance on a "notice-and-choice" framework, the field's emphasis on individual control, and a general lack of protection for data once it is publicly exposed, create conditions for collapse.[48] When leveraged by AI companies, these latent conditions further exacerbate the collapse of the regimes into one another. In other words, collapse exposes and intensifies privacy law's underlying structural weaknesses, creating a feedback loop that sophisticated actors can exploit—and, in turn, further exacerbating the collapse.

Critically, in the sense used in this Article, inter-regime doctrinal collapse is not itself good or bad. Collapse is a structural condition. The question is what conduct it facilitates or impedes. In some contexts, inter-regime doctrinal collapse enables positive outcomes. Collapse might produce useful blending of domains, in a way that makes overlapping regimes more coherent. Collapse might promote flexibility, adaptation, and rapid responsiveness to social and technological changes. And collapse might create space for better tailoring of legal arguments, in ways that more closely track the real-world impact of regulatory regimes, as opposed to insisting on legal formalism. Consider, for example, the Federal Trade Commission's privacy and data security enforcement

[46] See discussion of the distinct logics of copyright law and information privacy law *infra* Parts II.A–B.

[47] See discussion and sources cited *infra* note 107.

[48] See discussion *infra* Part II.

portfolio.[49] The agency's digital privacy work, beginning in the late 20th century, arguably represents a convergence of privacy law and consumer protection law, in adaptive, flexible, and entirely lawful ways.[50] Law has play in the joints. Perhaps collapse opens up new possibilities.

Whether collapse is good or bad, as a normative matter, is a contextual judgment that depends on the forms of legal and social exploitation that it enables and the real-world repercussions of that exploitation. Parts III and IV argue that the emerging patterns of exploitation in AI and data acquisition are likely to prioritize the interests of privileged, sophisticated private actors, at the expense of individuals and public legitimacy. That, I contend, is a net negative development. Even accepting that legal regimes will inevitably overlap to some extent, legal rules and doctrines appear arbitrary when they do not consistently allocate rights or constrain power. And even accepting that ambiguity in the law can promote flexibility and adaptation, law loses its form when it becomes infinitely "elastic."[51] Although legal systems can try to muddle through the collapse of fields that govern data, there are steep costs for individuals, for society, and for institutions.[52]

To be sure, information privacy law and law and technology scholars have previously raised questions about conceptual overlap and regulatory breakdown around data, as well as the political economy of data governance. For instance, Professor Viljoen critiques the individualistic orientation of current data governance laws. She contends that contemporary data governance fails to confront the collective and relational aspects of data production and results in systemic legal shortcomings.[53] Moreover, a rapidly growing literature analyzes not only AI and copyright battles, but also the ways that AI may put pressure on the underlying premises of copyright law itself.[54] In addition to doctrinal

[49] *See* Alicia Solow-Niederman, *The Overton Window and Privacy Enforcement*, 34 HARV. J. L. & TECH. 1007 (2024) (arguing that the FTC's unfair and deceptive acts and practices authority should be understood as a flexible, adaptable "Overton Window," and not a static jurisdictional charge).

[50] *See* Solow-Niederman, *The Overton Window and Privacy Enforcement*, *supra* note 49, at 1010–11 .

[51] *Cf.* Katharina Pistor, *Law's Elasticity: An Inquiry into the Relation of Law and Power in Finance*, EURO. J. SOC, 249, 250–55 (2021) (discussing "legal elasticity," power relations, and legal constraints in the context of financial systems).

[52] See discussion *infra* Part II.

[53] *See* Viljoen*, supra* note 26, at 609–13, 653–54.

[54] *See e.g.*, Peter Henderson & Mark A. Lemley, *The Mirage of Artificial Intelligence Terms of Use Restrictions*, 100 INDIANA L.J. 1327, 1335-36 (2025); Nancy S. Kim, *AI and the Fine Print Disruption of Copyright*, 129 PENN. ST. L. REV. 577, 580 (2025); B.J. Ard, *Copyright's Latent Space: Generative AI and the Limits of Fair Use*, 110 CORNELL L. REV. 509, 515 (2025); Blake E. Reid, *What Copyright Can't Do*, 52 PEPP. L. REV. 515, 529 (forthcoming 2025), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4766540; Robert Brauneis, *Copyright and the Training of Human Authors and Generative Machines*, 47 COLUM. J. L. &

arguments, some of this work has taken a theoretical and structural turn. Notably, Professor Goldenfein identifies contestation "over how to define the AI dataset as a legal and economic object" in the digital economy.[55] He argues that private actors switch between characterizing "datasets" as expressive "content" and as raw "data" to serve their bottom lines, "mak[ing] the simple application of copyright doctrine to AI training datasets deeply fraught."[56] I agree with Professor Viljoen that "[t]he unsettled status of data in law presents both a challenge and an opportunity,"[57] and I concur with Professor Goldenfein that AI development is producing legal instability.[58] But I contend that fully diagnosing the root cause requires digging deeper and focusing on the internal structures of information privacy law and copyright law, and the ways that these legal structures affect private power and public legitimacy.

This Article's focus on doctrinal collapse thus complements, but departs from, scholarship that focuses on law's complicity with dominant political and economic structures.[59] Critical scholars have long demonstrated how legal categories can entrench asymmetry, rationalize enclosure, and conceal power dynamics.[60] My analysis starts from a different place: Law's own formal (legal) and informal (social) doctrinal structures,[61] the internal tensions that emerge when multiple legal fields assert overlapping authority over the same regulatory

---

ARTS 1, 3 (forthcoming 2025); Carys J. Craig, *The AI-Copyright Trap*, 100 CHICAGO-KENT L. REV. 107, 108 (forthcoming 2025); Katherine Lee, A. Feder Cooper, & James Grimmelmann, *Talkin' 'Bout AI Generation: Copyright and the Generative AI Supply Chain*, 72 J. COPYRIGHT SOC'Y 251, 252 (2025); Niva Elkin-Koren, *Back to the Future: Navigating the Copyright/Contract Interface in the Generative AI Era*, 39 BERK. TECH. L.J. 1137, 1139 (2024); Mark Lemley, *How Generative AI Turns Copyright Law Upside Down*, 25 COLUM. SCI. & TECH. L. REV. 190, 195 (2024); Micaela Mantegna, *ARTificial: Why Copyright Is Not the Right Policy Tool to Deal with Generative AI*, 133 YALE L.J. FORUM 1126, 1128 (2024); Pamela Samuelson, *Fair Use Defenses in Disruptive Technology Cases,* 71 UCLA L. REV. 1484, 1547-48 (forthcoming 2024); Oren Bracha, *The Work of Copyright in the Age of Machine Production*, SSRN 1, 4 (Sep. 24, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4581738; Matthew Sag, *Copyright Safety for Generative AI*, 61 HOUS. L. REV. 295, 303 (2023). This body of work draws on a rich corpus of prior scholarship on the application of copyright law to automated systems and works authored by "machines" or "robots." To avoid making an already long footnote even longer, these works are not included here.

[55] Goldenfein, *supra* note 24, at 2.

[56] Goldenfein, *supra* note 24, at 3.

[57] Viljoen*, supra* note 26, at 654.

[58] *See* Goldenfein, *supra* note 24, at 2–3.

[59] *See, e.g.*, COHEN, BETWEEN TRUTH AND POWER, *supra* note 25; Amy Kapczynski, *The Law of Informational Capitalism*, 129 YALE L.J. 1460 (2020).

[60] *See, e.g.*, Kapczynski, *supra* note 59, at 1496-1514; Duncan Kennedy, *The Structure of Blackstone's Commentaries*, 28 BUFF. L. REV. 205, 211–17, 219–21 (1979).

[61] See discussion *infra* Part II.C (analyzing legal and social regulatory costs).

object, and the exploitation that this doctrinal collapse facilitates.[62] This starting point permits a more fine-grained analysis of the forms of behavior that collapse enables in the AI development context, who does and does not benefit, and how the legal structuring that produces collapse affects the rule of law itself.

B. Distinguishing Collapse from Other Regulatory Dynamics

Inter-regime doctrinal collapse matters because it destabilizes legal domains, thereby creating space for legal reconstruction. Because the overlap, blurring, and irreconcilable logics that feature in doctrinal collapse can enable downstream exploitation, it might seem similar to other regulatory dynamics, such as regulatory gaps or arbitrage. In general, limitations or uncertainties in the law often open the door to corporate manipulation. For instance, as Professor Cohen argues in her foundational work on zones of "legal immunity," regulatory frameworks can be constructed in ways that enable powerful entities in the digital economy to extract and control data without meaningful constraints or accountability.[63] And regulatory arbitrage is often strategic and exploitative (for better or for worse).[64] So, too, can doctrinal collapse permit private actors to strategically leverage the boundaries of legal domains.[65]

But it is a mistake to conflate collapse with other regulatory dynamics simply because doctrinal collapse can also lead to exploitation. Descriptively, collapse differs from regulatory gaps. Regulatory gaps involve the absence or insufficiency of law.[66] Inter-regime doctrinal collapse, in contrast, involves too much overlapping law. It is, moreover, descriptively and functionally different from regulatory arbitrage. Regulatory arbitrage refers broadly to the idea that a company can strategically arrange its operations to benefit from a particular set of legal or regulatory conditions.[67] More precisely, a private firm may be able to

---

[62] *Cf.* David Singh Grewal, Book Review, *The Laws of Capitalism*, 1128 HARV. L. REV. 626, 628–29 (2014) (reviewing Thomas Piketty, CAPITAL IN THE TWENTY-FIRST CENTURY (2014)) (arguing that the structure of law itself warrants analysis: "I develop an account of the 'laws' of capitalism, understood not as statistical regularities obtaining in a given socioeconomic regime, but as the legal structuring that undergirds it — in other words, the laws of capitalism understood as laws.").

[63] COHEN, BETWEEN TRUTH AND POWER, *supra* note 25 at 12, 97–100.

[64] García, *supra* note 67, at 203-04, 208 (citing Dan Burk, *Perverse Innovation*, 58 WILLIAM & MARY L. REV. 1, 15–18 (2016)) (analyzing exploitation of statutory language, characterizing regulatory arbitrage as "arguably exploitative," and acknowledging different perspectives on whether this is good or bad).

[65] See discussion of the political economy of doctrinal collapse *infra* Part III.

[66] *See* Rebecca Crootof & BJ Ard, *Structuring TechLaw*, 34 HARV. J. L. & TECH. 347, 360 (2021).

[67] There are many definitions of "regulatory arbitrage," but this sort of strategic manipulation is a common thread that runs through them. *See* Elizabeth Pollman, *Tech, Regulatory Arbitrage, and Limits*, 20 EUR. BUS. ORG. L. REV. 567, 571 (2019) ("Regulatory arbitrage has

make strategic choices to lower its regulatory costs and thereby reduce its overall costs.[68] Regulatory arbitrage is thus a regulatory maneuver. Collapse, in contrast, is not itself a strategic regulatory move; rather, it is the enabling condition for subsequent exploitation of both legal and social costs across the partially-overlapping domains.[69]

The next Part turns to the AI development context to illustrate collapse on the ground, assessing how data acquisition today is blurring formal legal lines and scrambling the irreconcilable normative logics that characterize copyright law and information privacy law.

II. Doctrinal Collapse on the Ground: AI and Data Acquisition

Doctrinal collapse is not theoretical or speculative; to the contrary, it is already happening. This Part analyzes contemporary AI development and data acquisition as a case study of collapse in action. AI development is a particularly salient example because it features a regulatory object (data) that is regulated by two fields of law (copyright law and information privacy law).[70] Each domain has historically operated with its own doctrinal rules, policy goals, and normative logic, as Part II.A assesses.

Moreover, collapse is occurring with unusual speed and urgency because a spate of AI litigation and regulation raises controversial and unsettled legal questions and comes with high economic and social stakes. Part III.B analyzes

---

been variously defined, but the term consistently includes the notion of manipulation or strategic design of an activity to take advantage of specific legal or regulatory treatment."); *see also* Kristelia A. García, *Copyright Arbitrage*, 107 CAL. L. REV. 199, 201 (2019). On regulatory arbitrage in the tax and corporate law setting, see, for example, Victor Fleischer, *Regulatory Arbitrage*, 89 TEX. L. REV. 227, 230 (2010); Frank Partnoy, *Financial Derivatives and the Costs of Regulatory Arbitrage*, 22 J. CORP. L. 211, 227 (1997).

[68] Pollman, *supra* note 67, at 571 ("Regulatory costs are engineered, not fixed or exogenous."). Private actors may lower regulatory costs in many ways. *Cf.* Annelise Riles, *Managing Regulatory Arbitrage: A Conflict of Laws Approach*, 47 CORNELL INT'L L. J. 63, 71 (2013) (identifying and defining jurisdictional arbitrage); Pollman, *supra* note 67, at 571 (defining categorical arbitrage) (citing Riles, *supra*, at 71).

[69] See discussion *infra* Part II.C.

[70] Technically speaking, copyright law is a sub-field of IP law. My analysis focuses on copyright law and information privacy law both because this relationship has been historically contested and because this interaction is where inter-regime doctrinal collapse is emerging with force in the context of data acquisition and AI development. Other scholars have considered forms of collapse and the blurring of boundaries within the field of IP law. *See generally* Bartholomew & Tehranian, *supra* note 30 (analyzing "borrowing" and "collapse" within IP law, with an eye to patent and copyright law); Geddes, *supra* note 30 (analyzing generative AI and the blurring of doctrines within IP law, with an eye to copyright and trademark law). My analysis is distinctive because it assesses copyright law and the substantively unrelated domain of information privacy law.

data acquisition and AI development and provides concrete examples of the collapse of copyright law and information privacy law. Part III.C then pinpoints two leading tactics—"buy" and "ask"—that private actors are using to engage in exploitation and leverage this collapse. Together, these parts provide a foundation for Part IV's analysis of political economy, institutional, and jurisprudential implications. Because doctrinal collapse and the exploitation that it enables are especially stark and significant in the AI context, this case study sets the stage for future work on similar dynamics that may emerge more subtly or slowly in other contexts.[71]

A. Distinct Logics

Scholars have long contested the boundaries between copyright law and information privacy law.[72] Even so, the two domains are distinct. Indeed, even scholarship contending that copyright law can be a tool to address privacy harms, such as the non-consensual dissemination of intimate imagery,[73] implicitly assumes that there *is* a difference between the two bodies of law. It would be difficult to use one domain to prevent and redress harms in the other unless the two regimes maintained discrete boundaries in the first place.

Copyright law and information privacy law are different in two central ways: First, they feature distinct doctrinal lines, in the sense that their controlling rules are different. Second, they feature distinct normative lines, in the sense that their goals are different.[74] Together, I refer to these doctrinal and normative considerations as the "animating logic" or "logic" of each regime.

An important caveat is in order: My discussion of each field's logic is necessarily descriptive and stylized. Descriptive, because I aim to represent the law as it presently is, in the American context. I do not argue that these logics

[71] This approach emulates the one I have taken in other work, both alone and with a co-author. *See* Alicia Solow-Niederman, *Information Privacy and the Inference Economy*, 117 NW. U. L. REV. 357, 363 & n.18 (2022) (arguing that machine learning and inference-derivation expose "weaknesses in information privacy law's current approach and . . . forecast emerging strains on its protective regime"); Richard M. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 STAN. TECH. L. REV. 242, 247 (2019) (offering that the study of AI judging "sheds light on governance issues that are likely to emerge more subtly or slowly elsewhere"). *See also* Aziz Z. Huq, *Constitutional Rights in the Machine-Learning State*, 105 CORNELL L. REV. 1875, 1879, 1885–86 (2020) (taking a similar approach).

[72] See sources cited *supra* note 19.

[73] *See, e.g.*, Bambauer, *supra* note 19, at 2030-31; Levendowski, *Using Copyright to Combat Revenge Porn*, *supra* note 19, at 439.

[74] *Cf.* Kaminski & Jones, *supra* note 2, at 1223 (developing a methodology of "legal construction of technology" and arguing that "[e]ach area of law has its driving theories, which typically prioritize particular values. These theories and their values aim, and indeed typically constrain, the law's construction." (internal citation omitted)).

reflect how copyright law or privacy law could or should be, nor do I make claims about how the fields should interact. Stylized, because I recognize that there is contestation within each of these fields of law[75] and I accept that other scholars might characterize the fields differently.[76] Nonetheless, I maintain that there is value in simplified, stylized understandings of the law.[77] I believe that the animating logics presented below are descriptively accurate of the dominant strands of American copyright and privacy law, as they currently are. This approach is useful: To assess whether collapse is a problem, and if so, why, we must engage with the doctrinal rules and the normative motivations that control the law as it presently is.

Start with the distinct doctrinal lines. This point is relatively straightforward: Different legal rules apply in copyright law, as compared to information privacy law. For example, although both domains involve "fairness" concerns, each domain has vastly different legal meanings and controlling legal rules. If a lawyer mixed up the "fair use" doctrine in IP law[78] and the "fair information

[75] For an example of fault lines within privacy law, compare, for instance, María P. Angel & Ryan Calo, *Distinguishing Privacy Law: A Critique of Privacy as Social Taxonomy*, 124 COLUM. L. REV. 507, 511, 552–53 (2024) (arguing "that the long-dominant social-taxonomic approach to privacy and privacy law is no longer serving the field" and advocating a "post-taxonomic approach to privacy" that critically assesses "the reasons why a given problem merits study under a privacy framework") with Daniel J. Solove, *Against Privacy Essentialism* (Dec. 9, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4826385 (manuscript at 5–6 and sources cited therein) (responding to Professors Angel and Calo: "'Privacy' is an umbrella term that refers to a group of related issues that are fruitful to discuss and address together. Instead of fixating on the meaning of the word 'privacy,' it is more productive to examine particular problems." (citations omitted)).

[76] Among many possible citations, see, for example, SILBEY, AGAINST PROGRESS, *supra* note 19, at 156–62 (arguing that "[c]opyright and privacy are especially fruitful collaborators because copyrighted expression embodied in a tangible object, such as a photograph, missive, or manuscript, often contains personal information" and recognizing that intellectual property interests and constitutional privacy "substantially overlap" at the same time that the goal of intellectual property law can conflict with privacy protections); SOLOVE, THE FUTURE OF REPUTATION, *supra* note 19, at 185 ("Copyright and privacy are both ways of controlling information." (citing Zittrain, *supra* note 19, at 1203 and Lawrence Lessig, *Privacy as Property*, 69 SOC. RESEARCH 247, 250 (2002))).

[77] *See* Re & Solow-Niederman, *supra* note 71, at 252 (presenting "two stylized models of adjudicatory justice," acknowledging that "these models simplify complex jurisprudential questions and processes," and contending that "delineating and contrasting these two views of adjudication" nonetheless reveals important interactions between AI and the juridical system). *Cf.* Mireille Hildebrandt, *Domains of Law: Private, Public, and Criminal Law*, *in* LAW FOR COMPUTER SCIENTISTS AND OTHER FOLK 39, 42 (2020) (identifying "three major domains: private, public and criminal law" and discussing "subdomains," such as "(1) constitutional law, (2) administrative law, [and] (3) international public law" for public law).

[78] *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 577–78 (1994); Harper & Row Publishers, Inc. v. Nation Enterprises, 471 U.S. 539, 560–61 (1985).

practice principles" in information privacy law,[79] that conflation of doctrines would be an obvious legal error. So, too, if a lawyer attempted to apply the Federal Trade Commission's unfairness standard,[80] which is often invoked in privacy enforcement actions by the agency,[81] rather than the fair use defense's multi-factor balancing test, which arises in private copyright litigation.[82] Because different rules apply in each domain, arguments such as these would be legally nonsensical and doctrinally incoherent.

The normative lines of copyright law and information privacy law also diverge.[83] Copyright law, as expressed in the dominant contemporary understanding in American law, focuses on how to generate incentives to create.[84] It calibrates these incentives through a system of property rights that assigns limited, transferable rights over intangible assets and thereby promotes creative production and dissemination.[85] As Jessica Litman summarizes, "[t]he

---

[79] *See* U.S. DEP'T OF HEALTH, EDUC. & WELFARE, RECORDS, COMPUTERS, AND THE RIGHTS OF CITIZENS 41–42 (1973) (describing "[s]afeguards for personal privacy . . . [that] would require adherence by record-keeping organizations to certain fundamental principles of fair information practice"); *id.* at 50 (recommending the enactment of "a Code of Fair Information Practice for all Automated personal data systems"). *See also* Woodrow Hartzog, *The Inadequate, Invaluable Fair Information Practices*, 76 MD. L. REV. 952, 952, 956–57 (2017) (discussing the FIPs and their proceduralized approach to privacy protection).

[80] *See* Letter from the FTC to Wendell Ford & John Danforth, S. Comm. on Com., Sci. & Transp., Commission Statement of Policy on the Scope of Consumer Unfairness Jurisdiction (December 17, 1980), reprinted in In re Int'l Harvester Co., 104 F.T.C. 949 app. at 1070–76 (1984) (clarifying the FTC's standard for unfairness); 15 U.S.C. § 45(n) (1994) (codification of standard articulated in FTC's 1980 Letter).

[81] *See* Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 599 (2014).

[82] *See, e.g.*, sources cited *supra* note 78.

[83] *See, e.g.*, Litman, *supra* note 19, at 1295–96 (rejecting a "property rights model" for data privacy protection as an "ineffective" approach that "would, in all likelihood, make the problem worse" and identifying "intellectual property [as] . . . the paradigmatic example" of a property model); Samuelson, *Privacy as Intellectual Property*, *supra* note 19, at 1129 (rejecting the propertization of data privacy because "[d]eep differences in the purposes and mechanisms of traditional intellectual property rights regimes and the proposed property rights regime in personal data raise serious doubts about the viability of a property rights approach and about its prospects of achieving information privacy goals"); Lemley, *Private Property*, *supra* note 19, at 1547 (concurring with Professor Samuelson and further contending that "creating an intellectual property right in individual data is a very bad idea").

[84] *See* Mark A. Lemley, *Property, Intellectual Property, and Free Riding*, 83 TEX. L. REV. 1031, 1031 (2005).

[85] *See, e.g.*, Wendy J. Gordon, *Fair Use as Market Failure: A Structural and Economic Analysis of the Betamax Case and its Predecessors*, 82 COLUM. L. REV. 1600, 1610 (1982) ("Copyright and patent law create ownership rights in intellectual property, with the primary goal of generating monetary incentives for the production of creative works, thereby 'promot[ing] the Progress of Science and useful Arts.'" (quoting U.S. Const. art. I, § 8, cl. 8)). *But see id.* at 1610 n.63 ("The role of copyright law in the maintenance of economic incentives has been a matter of significant debate.").

chief justification for so thoroughly commodifying rights in creative output is that it facilitates their transfer and exploitation."[86] The idea is that a system of "easy transfer" and "easy exploitation" permits entities to purchase and use copyright rights, and simultaneously "persuades authors and distributors to invest their resources in the creation and dissemination of works of authorship, while encouraging the widest profitable distribution of copyrighted works."[87] That said, even under this utilitarian view of copyright law,[88] the system does not grant unlimited property rights. The incentive structure that it creates is tightly linked to the importance of the public domain and the preservation of space for creative expression and public access to creative works.[89] A regime focused on properly calibrated authorial incentives is thus bedrock for contemporary copyright law. Indeed, even scholars skeptical of this incentive-oriented frame grant its dominance.[90]

In contrast to copyright law's focus on incentives, contemporary American information privacy law emphasizes a different concept: Control.[91] Specifically,

---

[86] Litman, *supra* note 19, at 1297 (citing PAUL GOLDSTEIN, COPYRIGHT: PRINCIPLES, LAW AND PRACTICE 3-11 (1989)). *See also* Fromer, *supra* note 19, at 551 ("Patent and copyright laws are designed to encourage the creation and dissemination of socially valuable works in their respective spheres: science and technology for patent, and arts and culture for copyright.").

[87] Litman, *supra* note 19, at 1297.

[88] *Id*.

[89] *See* Ard, *supra* note 54, at 572 ("Copyright protects copyright owners against intrusion upon authorial value. . . . However, . . . copyright law has historically allowed others to exploit a work's non-authorial value—that which flows from the work's non-original or non-expressive elements, its use of tropes that derive value from societal expectations rather than the author's creative choices, and in some instances from third-party contributions."). *Cf.* Robert Brauneis, *Copyright and the Training of Human Authors and Generative Machines*, 48 COLUM. J. L. & ARTS 1, 40 (2024) ("[T]he purpose of copyright law is not only to protect and incentivize the creation of aesthetic experiences—felt expression. It is also, and perhaps primarily, to protect and incentivize the creation of learning experiences.").

[90] *See, e.g.*, JULIE E. COHEN, CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE 61–62 (2012) ("Anglo-American copyright is premised on a set of assumptions about the relationship between copyright and creativity that most scholars largely accept: copyright supplies incentives for authors to produce creative work, but the creative process is essentially internal and unknowable. . . . This account of cultural development is incomplete in every critical respect."); Julie E. Cohen, *Copyright as Property in the Post-Industrial Economy: A Research Agenda*, 2011 WISC. L REV 141, 142–43 ("The statement that the purpose of copyright is to furnish incentives for authors has attained the status of a rote incantation. . . . The incentives-for-authors formulation of copyright's purpose is so deeply ingrained in our discourse and our thought processes that it is astonishingly hard to avoid invoking. . . ." (citation omitted)).

[91] *See, e.g.*, Julie E. Cohen, *Turning Privacy Inside Out*, 20 THEORETICAL INQUIRIES L. 1, 3 & n.3 (2019) ("Perhaps the dominant justification for privacy is that it promotes and protects individual autonomy." (citing BEATE RÖSSLER, THE VALUE OF PRIVACY (trans. R.D.V. Glasgow) (2d ed. 2018); Anita L. Allen, *Coercing Privacy*, 40 WM. & MARY L. REV. 723, 738–40 (1999))); ARI EZRA WALDMAN, PRIVACY AS TRUST: INFORMATION PRIVACY FOR AN INFORMATION AGE 29–33 (2018) (discussing literature that conceptualizes privacy as

the central regulatory paradigm, especially at the federal level, prioritizes giving individuals control over access to their own, personal information.[92] The primacy of control is typically closely connected to the importance of autonomy and a liberal conception of the self.[93]

This control-centered approach is especially evident in the "consumer protection" model of information privacy law that dominates at the federal level.[94] As William McGeveran has explained, this consumer protection approach "generally allows any collection and processing of personal data, unless it is specifically forbidden."[95] Central to this model is individual "'notice' of, and 'consent' to, [companies'] collection and use of their data."[96] In many areas, there are not overarching federal statutory protections for collection and processing of consumer data; rather, "sectoral" statutes apply only to specific categories of sensitive information (such as health data or financial data), and only where particular conditions are satisfied.[97] Where there are statutory protections, control remains a guiding principle: Privacy law generally relies on individual rights that are intended to "provide people with control over their personal data."[98]

"autonomy, choice, and control"); Paul M. Schwartz, *Privacy and Democracy in Cyberspace,* 52 VAND. L. REV. 1609, 1613 & n.15 (1999) (identifying "the traditional liberal understanding of information privacy, which views privacy as a right to control the use of one's personal data").

[92] On the centrality of control in privacy law, as a general matter, see sources cited *supra* note 91; *see also* Viljoen, *supra* note 26, at 598–600 (surveying "Traditional Accounts: Privacy as Control and Access"). For discussion of the centrality of control in federal information privacy law, see Solow-Niederman, *Information Privacy and the Inference Economy*, *supra* note 71, at 369–78, and sources cited therein; Viljoen, supra note 26, at 592–94.

[93] *But see* Julie E. Cohen, *What Privacy Is For*, 126 HARV. L. REV. 1904, 1905–06 (2013) (contesting the standard liberal account and arguing that "the self who is the real subject of privacy law and policy is socially constructed[.]")

[94] *See* William McGeveran, *Friending the Privacy Regulators*, 58 ARIZ. L. REV. 959, 973–79 (2016) (canvassing "[t]he American Consumer Protection Model" in privacy law).

[95] *Id.* at 966.

[96] Solow-Niederman, *Information Privacy and the Inference Economy*, *supra* note 71, at 370 (citing McGeveran, *supra* note 94, at 978).

[97] *See* Solow-Niederman, *Information Privacy and the Inference Economy*, *supra* note 71, at 370–71 & n.56, and sources cited therein.

[98] *See* Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1880 (2013). Again, privacy law scholars have long critiqued this model. On critiques of an individual-centered approach, see, for example, Viljoen, *supra* note 26, at 578–79; Karen Levy & Solon Barocas, *Privacy Dependencies*, 95 WASH. L. REV. 555, 557–58 (2020); Neil Richards & Woodrow Hartzog, *A Relational Turn for Data Protection?*, 6 EURO. DATA PROTECTION L. REV. 492, 493-94 & 494 n.9 (2020) (compiling privacy law scholarship focused on relationships). On critiques of a reliance on individual rights, see, for example, Daniel J. Solove, *The Limitations of Privacy Rights*, 98 NOTRE DAME L. REV. 975, 984–85 (2023); Julie E. Cohen, *How (Not) To Write a Privacy Law*, *in* DATA AND

Outside of these sectoral statutory protections, most data processing is governed by a company's privacy policy and terms of service (ToS), with the backstop of FTC regulatory enforcement.[99] Control is key here, too: Under the so-called "notice-and-choice" approach, the idea is to provide a consumer with control over their own data, by giving that person a statement of the company's privacy practices ("notice") and obtaining individual consent to those practices ("choice") in exchange for access to a good or service.[100]

Privacy's control logic cashes out with a series of associated presumptions. As one example, U.S. law tends to discount any privacy interest in information once it is exposed in public, resting at least in part on the premise that the person exercised control when they disclosed it and extinguished any further privacy interest.[101] The bottom line for information privacy law on the ground remains control.[102]

DEMOCRACY: KNIGHT FIRST AMEND. INST., COLUM. U. 3–4 (2021). *Cf.* Margot Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1578 (2019) (arguing that algorithmic accountability requires both individual rights and complementary, systemic forms of accountability).

[99] *See* Solow-Niederman, *Information Privacy and the Inference Economy*, *supra* note 71, at 370, 374-75.

[100] *See* Solow-Niederman, *Information Privacy and the Inference Economy*, *supra* note 71, at 370. Control remains central even when the FTC pursues an enforcement action. *See id.* at 374 ("[E]nforcement actions . . . reflect the same core calculation: the objective is to define privacy in terms of an individual's control over information about them, as expressed through the exercise of notice and consent rights.").

[101] This result is clearest in Fourth Amendment law, where the "third party doctrine" has long maintained that a privacy interest against the government is extinguished once information is shared with any other individual or entity. *See* Orin S. Kerr, *The Case for the Third-Party Doctrine*, 107 MICH. L. REV. 561, 563 (2009). Judicial interpretations in the civil sector tend to follow this same understanding. *See* DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 108 (8TH ED. 2024) (discussing publicity of private facts tort and noting that "many courts hold that matters cease to be 'private' when occurring in public"); Nissenbaum, sources cited *supra* note 102. The privacy torts are a partial exception, but their reach is limited. *See, e.g.*, Danielle Keats Citron, *Mainstreaming Privacy Torts*, 98 CAL. L. REV. 1805, 1825-28 (2010); Neil M. Richards & Daniel J. Solove, *Privacy's Other Path: Recovering the Law of Confidentiality*, 96 GEO. L. J. 123, 153–54 (2007). Even cutting-edge state privacy legislation generally does not apply to information that is "publicly available," on the grounds that it is not the kind of data worthy of protection. *See, e.g.*, California Consumer Privacy Act (CCPA) of 2018, CAL. CIV. CODE § 1798.140(v)(2)(A) (West); Colorado Privacy Act, § 6-1-1303(17), https://coag.gov/app/uploads/2022/01/SB-21-190-CPA_Final.pdf.

[102] This is a descriptive claim. At the risk of becoming a broken record, privacy law scholars have critiqued this model, at length. For critique of the notice-and choice approach, see, e.g., Joel R. Reidenberg et al., *Privacy Harms and the Effectiveness of the Notice and Choice Framework*, 11 I/S: J.L. & POL. INFO. SOC'Y 485, 490–96 & nn. 17-44 (2015) (analyzing and citing to capacious literature criticizing the notice and choice system). *But see* Michael Birnhack, *In Defense of Privacy-as-Control (Properly Understood)*, 65 JURIMETRICS J. 143 (2026) (manuscript at 2) (defending "privacy-as-control" and contending the critique . . . too

* * *

Structurally, IP law and information privacy law are distinct, both in terms of doctrine and in terms of normative goals. They involve different formal legal rules; moreover, they involve different underlying goals, as encoded in the dominant strands of American law, on the ground. Copyright law emphasizes incentives (and uses a property regime to calibrate them); privacy law emphasizes individual control (and often invokes personal autonomy as a reason for this approach). The animating logics of the two fields diverge.

The next Part analyzes real-world examples from AI development to reveal how, despite fundamentally different and often-irreconcilable animating logics, the boundaries between IP and information privacy are functionally blurring and converging.

B. Blurring of Domains

Data is at the heart of AI and doctrinal collapse. Because data is a fundamental building block for AI tools, developers need data—lots of it.[103] And both copyright law and information privacy law claim legal authority to regulate data. This Part illustrates how the demand for data to construct AI systems makes

often reduces "control" to notice-and-choice/consent"). For critique of the argument that there is no privacy in public, and arguments for a more nuanced approach, see, e.g., Helen Nissenbaum, *Protecting Privacy in an Information Age: The Problem of Privacy in Public*, 17 L. & PHIL. 559, 560 (1998; Helen Nissenbaum, *Privacy as Contextual Integrity*, 79 WASH. L. REV. 119, 136–38 (2004); HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE 125-26 (2010); Helen Nissenbaum, *A Contextual Approach to Privacy Online*, 140 DAEDALUS 32, 37 (2011); Helen Nissenbaum, *Contextual Integrity Up and Down the Data Food Chain*, 20 THEORETICAL INQUIRIES L. 221, 238 (2019).

[103] Today's AI systems are a form of so-called "machine learning," which is data hungry. *See, e.g.*, KATE CRAWFORD, ATLAS OF AI 96–97 (2021); Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GA. ST. U. L. REV. 1305, 1316 (2019); David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 678 (2017). To date, access to a great deal of data has been tied to continued AI progress. *See* Katherine Lee, A. Feder Cooper, James Grimmelman & Daphne Ippolito, *AI and Law: The Next Generation* (July 6, 2023) (manuscript at 5), https://afedercooper.info/paper/lee2023explainers.pdf. Technical developments might change the status quo. For an argument that "synthetic data" presents a solution that can "mitigate many of the technical and legal challenges of real-world data" see Peter Lee, *Synthetic Data and the Future of AI*, 110 CORN. L. REV. 1, 4–5 (2025). But even if such developments are in AI's future, it's fair to say that development today demands a great deal of data, that companies would not have been able to create today's leading systems without access to it, and that data remains an essential input for contemporary AI systems—creating incentives to acquire it, en masse. For an argument that building ever-larger models won't continue to produce AI gains, see Arvind Narayanan & Sayash Kapoor, *AI Scaling Myths*, AI SNAKE OIL (June 27, 2024), https://www.aisnakeoil.com/p/ai-scaling-myths.

longstanding doctrinal realities newly salient and produces a state of inter-regime doctrinal collapse.

In AI development, doctrinal collapse occurs when the lines between copyright law and privacy law overlap and blur, with irreconcilable claims across the two regimes. This phenomenon is clearest when companies make claims about the "public" nature of data to justify data acquisition (under both copyright law and privacy law), and then simultaneously or subsequently make incompatible claims that the data is proprietary (again, potentially invoking copyright law as well as privacy law).[104] These contentions combine two kinds of doctrinal collapse: One, collapse that stems from conceptual blurring, as the category of "public" loses meaning, and two, collapse that arises from functional blurring, as copyright law and privacy law fail to predictably and consistently guide or constrain conduct.[105] Together, corporate rhetoric and legal arguments exploit different understandings of "public" data, enabling companies to switch between legal regimes in ways that further destabilize each domain's doctrinal integrity and normative coherence.[106]

OpenAI provides an especially striking example of how companies offer initial claims that data are publicly available and free to appropriate and use, only to later pivot to claims that data are confidential and closed. These arguments sound in both copyright law and privacy law, wielding the "public" nature of the material to shield the company from legal liability or social exposure in either field of law. In pending lawsuits, the company has asserted in public-facing materials and in myriad briefs that "[t]raining AI models using publicly available internet materials is fair use."[107] This part of the argument is a copyright law

[104] Unless otherwise specified, this Article uses the terms "public" and "publicly accessible" to refer to data that it is possible to obtain through the open internet or through other public means. It uses the term "publicly available" in reference to statutes that use this term, *see* discussion *infra* note 101, or in reference to AI developers' own claims, *see* discussion *infra* text accompanying notes 107–112.

[105] Thank you to Ben Sobel for helpful comments about the relationship between rhetorical and legal arguments. On these two types of collapse, see discussion *supra* text accompanying notes 47–48.

[106] For an early analysis of privacy, copyright, and control over public information, see SOLOVE, THE FUTURE OF REPUTATION, *supra* note 19, at 185 (contending that copyright permits "strong rights of control even though information is public," whereas "[c]ontrol in the privacy context is seen as outlandish or impossible"). I agree with Professor Solove that both regimes share a common interest in regulating public information; my contention is that, even if one accepts that both regimes are interested in controlling information, there are different animating logics at the core of each domain. A control logic that focuses on creative incentives and property interests is quite different from one that focuses on autonomy interests.

[107] *Open AI and Journalism*, OPEN AI BLOG (Jan. 8, 2024), https://openai.com/blog/openai-and-journalism. The company is a defendant in a spate of lawsuits alleging that it has taken creative content and, in the process of training AI models and producing outputs, copied the material in ways that infringe copyright protections. In litigation, OpenAI has raised the affirmative defense of fair use for its use of publicly available material to train its AI models.

claim. If the affirmative defense of fair use applies, then the company cannot be found liable for copyright infringement for the use of works to train its models.[108] This would be the result, as a matter of current copyright doctrine, whether or not this outcome serves copyright law's animating normative logic by properly calibrating incentives.[109] In addition, the "public" nature of this data weakens privacy law concerns. Companies like OpenAI can rely on the doctrinal status quo: Because American privacy law has long discounted any privacy interest in data that is exposed in public,[110] the same "publicly available" claim insulates the company from potential privacy objections. This is the case, as a matter of current privacy law, whether or not this outcome serves privacy law's animating normative logic by permitting an individual to control access to information about them.[111] OpenAI's argument thus picks up on the idea of "public" versus proprietary (in copyright law) and "public" versus private (in privacy law).

See, e.g., Memorandum of Law in Support of OpenAI Defendants' Motion to Dismiss at 2–3, N.Y. Times v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. Feb. 26, 2024) (identifying the key issues as "whether it is fair use under copyright law to use publicly accessible content to train generative AI models to learn about language, grammar, and syntax, and to understand the facts that constitute humans' collective knowledge," and contending that "OpenAI and the other defendants in these lawsuits will ultimately prevail because no one . . . gets to monopolize facts or the rules of language" (internal citations omitted)); Reply Memorandum of Law in Further Support of OpenAI Defendants' Motion to Dismiss at 1–2, N.Y. Times v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. Mar. 18, 2024) ("[U]sing publicly available information to extract uncopyrightable ideas and facts about language and the world to create a large language model that powers transformative generative artificial intelligence is a quintessential fair use under longstanding copyright doctrine."). The suits against OpenAI were transferred for multi-district litigation in April 2025. *See* Transfer Order, In Re: OpenAI, Inc., Copyright Infringement Litigation, MDL No. 3143 (J.P.M.L. Apr. 3, 2025), https://www.jpml.uscourts.gov/sites/jpml/files/MDL-3143-Transfer_Order-3-25.pdf.
OpenAI has continued to make similar arguments since this transfer. *See, e.g.*, In re: OpenAI, Inc., Copyright Infringement Litigation, Case No. 1:25-md-03143-SHS-OTW, OpenAI's Answer to Daily News Plaintiffs' Complaint, Doc. No. 8 (Apr. 29, 2025), at 4 ("OpenAI admits that it believes that training AI models using publicly available internet materials is fair use, as supported by longstanding and widely accepted precedents.") In addition, there are many other pending cases against other AI developers that raise similar issues regarding whether the use of data to train a generative AI model is a fair use. For ongoing discussion and tracking of generative AI lawsuits, see generally CHATGPT IS EATING THE WORLD, https://chatgptiseatingtheworld.com/ (last visited Oct. 17, 2025) (providing coverage, including but not limited to cases involving fair use arguments); *see also Database of AI Litigation*, ETHICALTECH@GW, https://blogs.gwu.edu/law-eti/ai-litigation-database-search (providing searchable repository of AI litigation).

[108] *See Campbell*, *supra* note 78, at 575–78, 590 (describing the history and purpose of fair use, setting out the contemporary doctrine, its purpose, and analyzing it as an affirmative defense); *Harper & Row Publishers*, *supra* note 78, at 560–61 (discussing fair use and noting that "[t]he drafters [of the Copyright Act of 1976] . . . structured the provision as an affirmative defense requiring a case-by-case analysis.").

[109] See *supra* Part II.A for a discussion of copyright law's "animating logic."

[110] *See supra* note 101 and accompanying text.

[111] See *supra* Part II.A for a discussion of copyright law's "animating logic."

These sorts of arguments about the "public" nature of data represent conceptual and functional blurring of the regimes of copyright and privacy law. The two regimes overlap and each apply, at least in theory, to the same set of data; moreover, there's a blurring of privacy and copyright considerations. For one, because some of these claims are made in court, and others in public-facing rhetoric, it's not entirely clear which doctrinal lines the arguments actually reflect. For another, it's not clear how these arguments connect up to underlying normative positions in either copyright law or privacy law. An argument that information should be public because it creates the right incentive structure for creative production and dissemination is very different from an argument that information should be public because an individual lacks control in that data point. But in AI development, it's not apparent which one moves the argument, and why. With this overlap, blurring, and irreconcilability triad, the two domains have collapsed into one another.

Furthermore, the legal structure becomes further destabilized because OpenAI also draws on copyright and other areas of IP law both as a shield and as a sword. Sometimes, the company makes IP-maximizing arguments to shield disclosure of its own data. In litigation, arguments about proprietary information have featured in discovery disputes. For instance, in *Tremblay v. OpenAI*, OpenAI contended that the so-called "English Colang Dataset," which was used to train Chat-GPT4, should be shielded from production.[112] Among other arguments, OpenAI presented this material as a "proprietary dataset."[113] Moreover, as a general matter, the company has resisted calls to release its training data; though not explicit, part of the reason is ostensibly a concern with trade secrecy and protection of commercially valuable information. At other times, the company wields IP law's property-based regime as a sword. Consider OpenAI's public statement that the Chinese AI developer that created DeepSeek "may have inappropriately distilled our models."[114] Although OpenAI has not

[112] Joint Discovery Letter Brief at 3, Tremblay v. OpenAI, Inc., No. 3:23-cv-03223 (N.D. Cal Jan. 17, 2025), ECF No. 254 (presenting both parties' stances regarding what has been referred to as the "English Colang Dataset"); *see also AI Discovery Battles Heat up as AI Developer Ordered to Produce Training Data*, DEBEVOISE & PLIMPTON (Feb. 5, 2025), https://www.debevoise.com/-/media/files/insights/publications/2025/02/ai-discovery-battles-heat-up-as-ai-developer-order.pdf.

[113] Joint Discovery Letter Brief at 3, Tremblay v. OpenAI, Inc., *supra* note 112, at 3. Other AI companies have made similar arguments in discovery disputes. *See, e.g.*, Joint Discovery Dispute Statement Regarding Publishers' Challenges to Anthropic's Confidentiality Designations at 10, Concord Music Group, Inc. v. Anthropic PBC, No. 5:24-cv-03811 (N.D. Cal. May 29, 2025), ECF No. 380 (statement filed by Anthropic asserting that "[d]isclosing [s]pecific [t]raining [d]ataset [w]ould [c]ause [s]erious [c]ompetitive [h]arm").

[114] Cade Metz, *OpenAI Says DeepSeek May Have Improperly Harvested Its Data*, N.Y. TIMES (Jan. 29, 2025), https://www.nytimes.com/2025/01/29/technology/openai-deepseek-data-harvest.html*; see also* Cristina Criddle & Eleanor Olcott, *OpenAI Says It Has Evidence*

filed a lawsuit formalizing these allegations, nor articulated specific copyright claims, this rhetoric suggests that DeepSeek's use of ChatGPT outputs to develop its models amounts to improper copying of OpenAI's proprietary content.[115]

These additional, IP-maximizing arguments, plus the original set of copyright arguments about fair use, contrast with a paucity of privacy law contentions.[116] Information privacy law is, in theory, germane for an AI company. Return, once more, to the example of OpenAI: As the company's own privacy policy acknowledges, OpenAI processes personal data, including user prompts, uploaded files, and other interactions.[117] In the United States, unless a user explicitly opts out, this user-provided content is used to train OpenAI's models.[118] Particularly as the company adds enhanced voice functionality and moves into "agentic" AI that can execute discrete tasks (like purchasing a product or cancelling a service) on a user's behalf,[119] at least some of this data

China's Deepseek Used Its Model to Train Competitor (Jan. 28, 2025), https://www.ft.com/content/a0dfedd1-5255-4fa9-8ccc-1fe01de87ea6.

[115] There is irony in this argument, as others have observed. *See* Chris Smith, *OpenAI Says It Has Evidence Deepseek Used ChatGPT to Train Its AI*, BGR (Jan. 29. 2025, 6:50 AM), https://bgr.com/tech/openai-says-it-has-evidence-deepseek-used-chatgpt-to-train-its-ai/ ("Ironically, if OpenAI's claim is true, it'll make the company experience what many creators felt when they discovered OpenAI may have trained its ChatGPT models using copyrighted materials without consent.").

[116] Other scholars have begun to observe this differential coverage. *See* Thomas D. Haley, *The Second Life of Information*, FLA. L. REV. (forthcoming 2025) (manuscript at 38) (manuscript on file with author) (contending that "[c]opyright would seem to pose more of an issue" for AI development than privacy law).

[117] For the privacy policy that applies to American users, see *Privacy Policy*, OPENAI, https://openai.com/policies/row-privacy-policy/ ("User Content: We collect Personal Data that you provide in the input to our Services ('Content'), including your prompts and other content you upload, such as files, images, and audio, depending on the features you use."). For the policy that applies for business and enterprise users, see *How Your Data Is Used to Improve Model Performance*, OPENAI, https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance.

[118] Individual users are given the choice to opt-out, albeit with leading language warning that they won't be helping to "improve the model for everyone." *Data Controls FAQ*, OPENAI, https://help.openai.com/en/articles/7730893-data-controls-faq (last visited Dec. 9, 2025); *see Consumer Privacy at OpenAI*, OPENAI (June 12, 2024), https://openai.com/consumer-privacy/.

[119] There is no settled definition of "agentic AI," and many companies invoke this term without precision. This Article follows the approach taken by a team of computer scientists at Princeton University, who consider AI systems "'agentic'" if they can pursue difficult goals without being instructed in complex environments[;] . . . if they can be instructed in natural language and act autonomously without supervision[;] [a]nd [if they] . . . are able to use tools, such as web search or programming, or are capable of planning." Melissa Heikkilä, *What Are AI Agents?*, MIT TECH REV. (July 4, 2024), https://www.technologyreview.com/2024/07/05/1094711/what-are-ai-agents/ (discussing Sayash Kapoor, Benedikt Stroebl, Zachary S. Siegel, Nitya Nadgir, Arvind Narayanan, *AI*

will be extremely personal or otherwise sensitive for the user. Thus, the control logic of privacy arguably applies to any such personal data processed by OpenAI.[120]

But even when privacy law's control logic is in theory on point, privacy is generally not the starting point for AI developers. In practice, American privacy law's contemporary treatment of what is "public" versus "private" has stripped privacy law of operational relevance.[121] AI developers tend not to make arguments about privacy that go beyond pointing out that information used for training is "public." User privacy might arise during discovery disputes—but only if doing so helps the company.[122] Instead, copyright law claims focused on "public" information dominate, without considering whether the results amply support the animating logics of both domains. When an entire domain that is normatively applicable can be swept aside or invoked only when it is useful to the company as a litigation tactic, it is a signal of collapse. The concern is not only that some values will receive short shrift, but also that one domain will shrink and another will expand in ways that distort the internal doctrinal structure of both legal regimes. Moreover, as the next Part discusses, collapse creates conditions for corporate manipulation of the two domains.

C. Doctrinal Instability, Regulatory Costs, and Exploitation

Although inter-regime doctrinal collapse is a neutral, structural condition, it has normative implications because of what it enables.[123] On the ground, collapse matters because it creates conditions in which private actors can convert doctrinal overlap and blurring into regulatory advantage—for better or for worse.[124] This Part again focuses on the AI context to detail the forms of

---

Agents That Matter, ARXIV (July 2, 2024), https://arxiv.org/pdf/2407.01502). *See also* Helen Toner et al., *Through the Chat Window and Into the Real World: Preparing for AI Agents*, CSET, at 4–7 (2024) (similarly describing "a cluster of properties that can be present to greater or lesser degrees, which together determine how 'agentic' an AI system is").

[120] This suggestion is not outlandish, as illustrated by the European Data Protection Board's guidance on the processing of personal data and AI models. *See* Eur. Data Prot. Bd., *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*, at 2 (Dec. 17, 2024), https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf ("For an AI model to be considered anonymous, both (1) the likelihood of direct (including probabilistic) extraction of personal data regarding individuals whose personal data were used to develop the model and (2) the likelihood of obtaining, intentionally or not, such personal data from queries, should be insignificant . . . .").

[121] For a discussion of privacy law's treatment of public data, see *supra* note 101 and accompanying text .

[122] For further discussion of discovery disputes, see *infra* Part II.B.

[123] *See* discussion *supra* Part II.

[124] *See* discussion *supra* Part II.

exploitation that collapse facilitates, thereby revealing regulatory dynamics that other accounts miss.

In general, inter-regime doctrinal collapse enables a particular form of exploitation in which a firm attempts to minimize its overall regulatory burdens across two overlapping legal regimes.[125] The relevant regulatory burdens include both formal legal constraints and incentives and informal social constraints and incentives. Together, these legal and social forces generate a set of regulatory costs for a private actor that is subject to both domains.

To make this point about cumulative regulatory burdens more concrete, picture a company that seeks to acquire data to train an AI model. The company may seek to reduce the overall legal and social costs that it bears for data acquisition, taking into account both information privacy law and copyright law. Such inter-regime cost-reduction is similar to arbitrage in that it seeks to reduce regulatory costs,[126] yet it is a distinct form of regulatory exploitation. The AI firm is not moving between distinct domains. It is not as if the AI developer says, "I wish to move out of the mainland of copyright and reside on the island of privacy," as it would in so-called "jurisdictional" regulatory arbitrage.[127] Nor does the firm assert, "My conduct removes my business from the statutory coverage of privacy law and should be understood to trigger copyright law obligations," as it would in so-called "categorical" regulatory arbitrage.[128] In conditions of doctrinal collapse, both fields of law continue to govern the same regulated object, and a firm that faces comparatively lower regulatory costs in one domain may seek to structure its affairs to minimize overall costs.[129] Regulation of data is a case in point: Both copyright law and privacy law regulate data, and companies face higher cumulative legal and social costs for data

[125] This use of the term "regulatory" does not refer to formal, top-down regulation that is promulgated by an administrative agency. I adopt the understanding of early internet law scholars and use the term "regulatory" to reference constraints and affordances that emerge from design choices, social norms, and market interactions, and not merely from formal law. *See* LAWRENCE LESSIG, CODE: VERSION 2.0, at 5–7 (2006); Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553, 554–55 (1998); *see also* James Grimmelmann, *Note, Regulation by Software*, 114 YALE L.J. 1719, 1722–23 (2005) (analyzing how software can serve as a regulatory modality). This Article uses the term "public regulation" to refer to top-down regulation by the state.

[126] For discussion of regulatory arbitrage and regulatory costs, see *supra* note 68 and accompanying text.

[127] Jurisdictional regulatory arbitrage occurs when a company takes advantage of different laws that apply in different jurisdictions. *See* Riles, *supra* note 68, at 71.

[128] Categorical regulatory arbitrage occurs when there are two functionally similar kinds of conduct or products and there is a legal or regulatory discrepancy between how the two are treated. *See* Pollman, *supra* note 68, at 571 (citing Riles, *supra* note 68, at 71).

[129] *See* discussion *supra* Part I.B (distinguishing arbitrage from collapse). Thank you to Chris Morten and Julie Cohen for helpful conversations that informed my thinking on this point.

acquisition in copyright law, as compared to cumulative legal and social costs for data acquisition in privacy law.

Start with the potential legal costs, which are straightforward. Statutes and public regulations create regulatory costs within a particular domain of law,[130] both through *ex ante* interventions that lead an entity to change its business operations to avoid legal liability, at a higher cost, and through *ex post* interventions like litigation and settlement expenses, damages awards, and regulatory penalties.

For AI companies, there are high legal costs imposed by copyright law, as compared to minimal legal costs imposed by privacy law. Notably, there are substantial potential costs from a wave of pending generative AI copyright litigation. There is a whole lot of potential legal liability and a whole lot of money involved in these cases.[131] Although some plaintiffs have abandoned their claims without judicial resolution,[132] prolonged litigation or expensive settlement arrangements negotiated after costly initial proceedings seem likely in most cases. For example, both the *New York Times* (a plaintiff) and OpenAI (a defendant) are well-resourced actors whose business models depend, respectively, on the ability to make a profit from producing news and on the ability to make a profit from producing AI models. When the *New York Times* alleges that OpenAI's "unlawful use of The Times's work to create artificial intelligence products that compete with it threatens The Times's ability to provide that service,"[133] each side is highly motivated to obtain a favorable legal resolution. Furthermore, with new lawsuits continuing to emerge, including

[130] *See* discussion *supra* note 68 and sources cited therein.

[131] Pamela Samuelson, *Legally Speaking How to Think about Remedies in the Generative AI Copyright Cases*, 67 COMMC'NS ACM, July 2024, at 29 ("If the plaintiffs succeed in claiming the uses of works as training data infringe copyrights, copyright statutory damage awards would almost certainly be staggeringly large . . . [and] could bankrupt most generative AI companies."). This prediction of eye-wateringly high damages and settlements, at a level that affects business' financial health, is playing out in real time. *See* Cristina Criddle & Lee Harris, *Insurers Balk at Multibillion-Dollar Claims Faced by OpenAI And Anthropic*, FIN. TIMES (Oct. 8, 2025), https://www.ft.com/content/0211e603-7da6-45a7-909a-96ec28bf6c5a?syn-25a6b1a6= ("OpenAI and Anthropic are considering using investor funds to settle potential claims from multibillion-dollar lawsuits, as insurers balk at providing comprehensive coverage for the risks associated with artificial intelligence."); Cade Metz, *Anthropic Agrees to Pay $1.5 Billion to Settle Lawsuit With Book Authors*, N.Y. TIMES (Sept. 5, 2025), https://www.nytimes.com/2025/09/05/technology/anthropic-settlement-copyright-ai.html (describing Anthropic's "landmark settlement" of $1.5 billion and reporting that the settlement, if approved by the court, will be "the largest payout in the history of U.S. copyright cases").

[132] *See, e.g.*, Notice of Voluntary Dismissal Without Prejudice at 1, P.M. v. OpenAI LP, No. 3:23-cv-03199 (N.D. Cal. Sep. 15, 2023).

[133] Complaint at 2, N.Y. Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. Dec. 27, 2023).

lawsuits that involve other technology companies[134] and startups[135] as well as lawsuits that involve leading AI developers like Google, Meta, and OpenAI, the legal battles are likely to persist. That may be good for lawyers, but it is not optimal for managing business expenses.

In contrast, the legal costs from the privacy law domain are minimal. Two early generative AI lawsuits, *Cousart v. OpenAI*[136] and *J.L v. Alphabet*,[137] are illustrative. Both lawsuits initially alleged that data scraping violated plaintiffs' privacy.[138] The privacy law claims didn't get far. The *Cousart* plaintiffs decided not to pursue any of their claims after the court dismissed the initial complaint,[139] and the *J.L.* plaintiffs filed an amended complaint that eliminates all privacy-related causes of action and contains a single copyright allegation.[140] There's much that could be said about how courts adjudicate privacy claims and undervalue privacy harms in assessing both justiciability and merits.[141] So, too,

---

[134] *See* Ashley Belanger, *Nvidia Sued Over AI Training Data as Copyright Clashes Continue*, ARS TECHNICA (March 11, 2024, 9:35 AM), https://arstechnica.com/tech-policy/2024/03/novelists-sue-nvidia-to-stop-spread-of-ai-models-trained-on-copyrighted-books/ (describing lawsuit against the chip maker Nvidia).

[135] *See* Kristin Robinson, *Major Labels Sue AI Firms Suno and Udio for Alleged Copyright Infringement*, BILLBOARD (June 26, 2024), https://www.billboard.com/pro/major-label-lawsuit-ai-firms-suno-udio-copyright-infringement/ (describing lawsuit against two AI music start-ups).

[136] Cousart v. OpenAI LP, No. 23-cv-04557-VC (N.D. Cal. 2024).

[137] J.L. v. Alphabet Inc., No. 3:23-cv-03440-AMO (N.D. Cal. 2023).

[138] Complaint at 105–111, A.T. v. OpenAI LP, No. 3:23-cv-04557-JCS (N.D. Cal. Sep. 5, 2023); Complaint at 71–73, J.L. v. Alphabet Inc., No. 3:23-cv-03440-AMO (N.D. Cal. July 11, 2023). *A.T. v. OpenAI* was subsequently recaptioned as *Cousart v. OpenAI; see A.T.*, *supra*, Order, ECF No. 77.

[139] Cousart v. OpenAI LP, No. 23-cv-04557-VC, 2024 WL 3282522, at *1 (N.D. Cal. May 24, 2024) (granting OpenAI's motion to dismiss). The judge in *Cousart* initially dismissed the complaint with leave to amend, chastising the plaintiffs for their "unnecessary and distracting" pleading (which ran to over 200 pages). Order Granting Motions to Dismiss, Cousart v. OpenAI LP, No. 3:23-cv-04557-VC (N.D. Cal. May 24, 2024). The *Cousart* plaintiffs filed notice that they would not submit a second amended complaint and requested that the court close the matter. Plaintiff's Notice of Intent Not to Amend First Amended Complaint at 1, Cousart v. OpenAI LP, No. 3:23-cv-04557-VC (N.D. Cal. June 14, 2024).

[140] The judge in *J.L.* dismissed the first amended complaint on similar grounds to the *Cousart* court, granting leave to amend. Order Granting Motion to Dismiss with Leave to Amend and Denying Administrative Motion to Relate Without Prejudice, J.L. v. Alphabet Inc., No. 3:23-cv-03440-AMO (N.D. Cal. June 6, 2024). Plaintiffs then filed a second amended complaint that removed the privacy allegations. *Compare* First Amended Complaint at 68–89, 123–25, Leovy v. Google LLC, No. 3:23-cv-03440-AMO (N.D. Cal. Jan. 5, 2024) (alleging violations of property, privacy, and copyright law and pressing two specific privacy claims) *with* Second Amended Complaint at 27, Leovy v. Google LLC, No. 3:23-cv-03440-AMO (N.D. Cal. June 27, 2024) (including one count alleging direct copyright infringement).

[141] *See, e.g.*, COHEN, BETWEEN TRUTH AND POWER, *supra* note 25, at 147. *See generally* Danielle Keats Citron & Daniel J. Solove, *Privacy Harms*, 102 B.U. L. REV. 793, 796-99

is there space to critique the sprawling initial complaints in these cases. But the practical upshot is clear: In the form it takes in most cases, privacy law imposes far lower legal costs than copyright law.[142]

In addition to these formal legal costs, regulation of private actors entails a set of informal social constraints. As Hillary Sale explains, the sociological concept of "social license" affects how firms operate.[143] Firms "exist with permission from the communities in which they are located, as well as with permission from the greater community and outside stakeholders."[144] Obtaining and maintaining social license requires firms to make investments to obtain public legitimacy and public trust.[145] Thus, social constraints impose regulatory costs on companies, too.

AI developers face steep social costs in copyright law and minimal social costs in privacy law. Copyright law features longstanding debates about the equity of "free riding" on one person's creation, including in ways that may displace their labor in the long run.[146] This issue arises with force in the generative AI context because the data that AI developers have scraped from the

(2022) (contending that courts have failed to recognize privacy harms and enumerating different privacy harms and their impacts).

[142] I refer to "most cases" because it is possible that other privacy-centered causes of action brought against other AI companies, predicated on different, allegedly privacy-invasive facts, could alter the status quo. For instance, in November 2025, plaintiffs filed a case alleging that "Google secretly turned on Gemini for all its users Gmail, Chat, and Meet accounts, enabling AI to track its users' private communications . . . without the users' knowledge or consent" and raising several privacy-focused legal claims. *See* First Amended Class Action Complaint at 2, 9–17, Thele v. Google LLC, Docket No. 5:25-cv-09704 (N.D. Cal. Nov 11, 2025). *But see* https://x.com/gmail/status/1991989459097653419 (Nov. 21, 2025), X post with public response from Google: "We do not use your Gmail content to train our Gemini AI model."). *Thele* is pending as of this writing. But regardless of the result there, cases like these remain far rarer than the myriad copyright-centered generative AI cases. That disparity remains, moreover, even if one considers privacy-focused lawsuits against more specialized AI providers, such as speech transcription services, to be in scope. *See, e.g.*, Consolidated Class Action Complaint, In re Otter.AI Privacy Litigation, Docket No. 5:25-cv-06911 (N.D. Cal. Aug 15, 2025) (privacy-focused lawsuit against Otter AI). There is much, much more copyright litigation. The upshot is simple: At least for now, the legal costs in copyright law remain far higher than those in privacy law.

[143] Hillary A. Sale, *The Corporate Purpose of Social License*, 94 S. CAL. L. REV. 785, 788 (2019).

[144] Sale, *supra* note 143, at 789.

[145] Sale, *supra* note 143, at 789–90.

[146] It may also move some courts. *See* Bryan Casey & Mark Lemley, *Fair Learning*, 99 TEX. L. REV. 743, 765-66 (2019) ("[C]ourts may well let their view of the equities creep into the analysis of the fourth [fair use] factor.").

internet includes the contributions of many content creators.[147] Although legal scholars have argued that it's a mistake to resolve these disputes through copyright doctrine,[148] the perception of inequitable appropriation remains powerful. For instance, authors have decried the use of their work as theft;[149] celebrities and artists have made prominent statements concerning AI labor appropriation;[150] and Hollywood actors have contested the use of generative AI and its potential effect on actors.[151] This is not mere rhetoric: In fall 2025,

---

[147] *See, e.g.*, Lee et. al, *Talkin' 'Bout AI Generation, supra* note 54, at 290, 314; Alice Xiang, *Fairness & Privacy in an Age of Generative AI*, 25 COLUM. SCI. & TECH. L. REV. 288, 304 (2024); Lee et. al., *AI and Law: The Next Generation*, *supra* note 103, at 10.

[148] *See, e.g.*, Jessica Silbey, *How Theories of Art Can Inform Our Debates about AI*, 74 EMORY L. REV. 1231, 1243–46 (2025); Ard, *supra* note 54, at 583–84; Reid, *supra* note 54, at 529–48; Mantegna*, supra* note 54, at 1158). *But see* Xiyin Tang, *Intellectual Property Law as Labor Policy*, 100 N.Y.U. L. REV. 62, 64, 67–68 (2025) ("This Article proposes a new, alternate framework; one that reveals how intellectual property has also functioned as labor policy—as a contested site through which creative laborers exchange work for wages and large IP firms amass power and capital, often at the expense of those laborers.").

[149] *See, e.g.*, Andrea Bartz, *I Sued Anthropic, and the Unthinkable Happened*, N.Y. TIMES (Sept. 29, 2025), https://www.nytimes.com/2025/09/29/opinion/anthropic-chatbot-lawsuit-books.html (guest essay by novelist and named plaintiff in *Bartz*) ("In August 2023, alone in my studio apartment, I learned that my most precious possessions had been stolen. . . . My heart raced as I typed my name into a database of works used to train large language models . . . Horror flooded my chest when several of my thrillers appeared.").

[150] *See, e.g.*, Dan Milmo, *Thom Yorke and Julianne Moore Join Thousands of Creatives in AI Warning*, THE GUARDIAN (Oct. 22, 2024, 1:49 PM), https://www.theguardian.com/film/2024/oct/22/thom-yorke-and-julianne-moore-join-thousands-of-creatives-in-ai-warning; Maria Sherman, *Miranda Lambert, Billie Eilish, Nicki Minaj Submit Letter to AI Developers to Honor Artists' Rights*, AP (April 2, 2024, 1:41 PM), https://apnews.com/article/ai-open-letter-billie-eilish-miranda-lambert-nicki-minaj-9cd5f32f692d83e75b9c3b3da1554b6f.

[151] *See* Kevin Collier, *Actors vs. AI: Strike Brings Focus to Emerging Use of Advanced Tech*, NBC NEWS (updated July 14, 2023, 5:14 PM), https://www.nbcnews.com/tech/tech-news/hollywood-actor-sag-aftra-ai-artificial-intelligence-strike-rcna94191. In addition to social contestation during the 2023 Hollywood strike, musicians such as Kate Bush and Paul McCartney and actors such as Julianne Moore and Kevin Bacon have called on the U.K. government "to protect artists from AI using their copyrighted works." *See infra* text accompanying note 150 (discussing public controversy in U.S.); *Kate Bush Joins Campaign Against AI Using Artists' Work Without Permission*, GUARDIAN (Dec. 12, 2024, 5:18 AM), https://www.theguardian.com/technology/2024/dec/12/kate-bush-joins-campaign-against-ai-using-artists-work-without-permission (discussing public controversy in U.K.). Moreover, there have been high-profile controversies involving celebrities, such as the contestation between the actress Scarlett Johansson and OpenAI over alleged emulation of her voice without permission. For a summary of this controversy, including arguments on both sides, see Nitasha Tiku, *OpenAI Didn't Copy Scarlett Johansson's Voice for ChatGPT, Records Show*, WASH. POST (updated May 23, 2024), https://www.washingtonpost.com/technology/2024/05/22/openai-scarlett-johansson-chatgpt-ai-voice/. This Article reserves concerns about appropriation law and the right to publicity to focus on distinct concerns about data scraping. On appropriation law and the right to publicity's potential application to generative AI, see Jennifer E. Rothman, Comment Letter

OpenAI introduced an "opt-out" default for copyright rights holders who did not wish to be included in the outputs of its Sora 2 text-to-video generation model.[152] After widespread protest, including public outcry from corporate rightsholders and industry trade groups, OpenAI shifted gears.[153] CEO Sam Altman promised to offer "more granular control" to rightsholders and sought patience from the public: "Please expect a very high rate of change from us; it reminds me of the early days of ChatGPT. We will make some good decisions and some missteps, but we will take feedback and try to fix the missteps very quickly."[154] Putting to the side whether Silicon Valley's "move fast and break things" mentality[155] is socially responsible or whether OpenAI's move was an intentional one to attract attention for Sora 2's launch,[156] the instant point is a different one: The social costs in copyright law and copyright-adjacent terrain are high, the controversy is a live one, and AI developers are, at the end of the day, concerned with community and stakeholder perceptions—so much so that social costs can, and indeed already have, shifted business practices.

To date, privacy law has not generated parallel social costs. It is generally believed that AI companies scraped the internet to acquire enough data to develop their models.[157] Such widespread scraping might be thought to raise

---

on Artificial Intelligence, Copyright, and Right of Publicity (Oct. 25, 2023), https://www.regulations.gov/comment/COLC-2023-0006-8229.

[152] *See* Keach Hagey, Berber Jin, & Ben Fritz, *OpenAI's New Sora Video Generator to Require Copyright Holders to Opt Out*, WALL ST. J. (Sept. 29, 2025, 6:36 PM), https://www.wsj.com/tech/ai/openais-new-sora-video-generator-to-require-copyright-holders-to-opt-out-071d8b2a?.

[153] *See* Jaures Yip, *OpenAI's Sora 2 Must Stop Allowing Copyright Infringement, Motion Picture Association Says*, CNBC (Oct. 7, 2025), https://www.cnbc.com/2025/10/07/openais-sora-2-must-stop-allowing-copyright-infringement-mpa-says.html. As others have noted, this shift to an opt-in policy only appears to cover outputs of the Sora 2 model; it does not speak to OpenAI's own use of copyright-protected data as inputs to train the model. *See* Dale Nelson & Phoenix Silkensen, *Sora 2 Does a Copyright Somersault Upon Launch*, FORBES (Oct. 17, 2025) https://www.forbes.com/sites/legalentertainment/2025/10/17/sora-2-does-a-copyright-somersault-upon-launch/.

[154] Facebook's internal motto in its early days was "move fast and break things." *See* Seth Fiegerman, *Are Facebook's 'Move Fast and Break Things' Days Over?*, MASHABLE (Mar. 13, 2014), https://mashable.com/2014/03/13/facebook-move-fast-break-things.

[155] Sora Update #1, SAM ALTMAN, https://blog.samaltman.com/sora-update-number-1 (Oct. 3, 2025, 8:37 PM).

[156] *See, e.g.*, *Furor Over Sora "Opt Out" For Copyright Owners. Is Sam Altman Gaslighting Again?,* CHATGPT IS EATING THE WORLD (Oct. 3, 2025), https://chatgptiseatingtheworld.com/2025/10/03/furor-over-sora-opt-out-for-copyright-owners-is-sam-altman-gaslighting-again/ (suggesting that OpenAI's Sam Altman is intentionally courting controversy to garner media attention for new product launches).

[157] *See* Sophie Bushwick, Lauren Leffer, Tulika Bose & Elah Feder, *Generative AI Models Are Sucking up Data from All over the Internet, Yours Included*, SCI. AM. (Oct. 23, 2023), https://www.scientificamerican.com/podcast/episode/generative-ai-models-are-sucking-data-up-from-all-over-the-internet-yours-included/. As of this writing, OpenAI does not disclose the sources of its training data. *Id.* The same is true for other leading generative-AI companies.

social costs, insofar as it contravenes social norms and violates the non-binding "robots.txt" internet protocol that website developers may use to signal that part or all of a website should not be scraped.[158] Indeed, the social costs might, in theory, seem substantial, particularly given public reporting that AI developers are routinely flouting social norms and technical protocols meant to shield at least some online data.[159] What's more, there is mounting scholarly debate about how to square mass scraping of publicly accessible data with protection of individuals' privacy interests.[160] But AI developers don't seem to have incurred social costs that are steep enough to compromise their social license. If they had, then they would ostensibly change their business practices. Instead, there has been relatively subdued privacy furor—at least to date—and very little about AI developers' business operations has changed. The AI startup Perplexity, for example, was involved in a scraping controversy[161] that led to an investigation by Amazon Web Services.[162] Yet the company remains a meaningful player in the AI sector, raising hundreds of millions of dollars from investors whose ranks include Amazon founder Jeff Bezos.[163] Perhaps monetary valuations like this do

See David Gray Widder, Meredith Whittaker & Sarah Myers West, *Why 'Open' AI Systems are Actually Closed, and Why This Matters*, 635 NATURE 827, 829 (2025).

[158] *See* Aaron Mak, *How to Stop AI from Eating the Open Internet*, POLITICO (Dec. 2, 2025, 5:14 PM), https://www.politico.com/newsletters/digital-future-daily/2025/12/02/how-to-stop-ai-from-eating-the-open-internet-00673326 (discussing robots.txt protocol); Jonathan Gillham, *Block AI Bots from Crawling Websites Using Robots.txt*, ORIGINALITY.AI (Aug. 22, 2024), https://originality.ai/ai-bot-blocking (finding, as of August 2024, that over one-third of the top 1000 websites were attempting to use the robots.txt signal to tell the OpenAI bot not to scrape their sites).

[159] *See, e.g.*, Katie Paul, *Exclusive: Multiple AI Companies Bypassing Web Standard to Scrape Publisher Sites, Licensing Firm Says*, REUTERS (June 21, 2024, 10:32 AM), https://www.reuters.com/technology/artificial-intelligence/multiple-ai-companies-bypassing-web-standard-scrape-publisher-sites-licensing-2024-06-21/; Kali Hays, *OpenAI and Anthropic Are Ignoring an Established Rule that Prevents Bots Scraping Online Content*, BUS. INSIDER (June 21, 2024, 3:04 PM PT), https://www.businessinsider.com/openai-anthropic-ai-ignore-rule-scraping-web-contect-robotstxt (reporting that OpenAI and Anthropic have ignored robots.txt protocol). *See also* Ina Fried, *For AI Firms, Anything "Public" Is Fair Game*, AXIOS (Apr. 5, 2024), https://www.axios.com/2024/04/05/open-ai-training-data-public-available-meaning (discussing, then critiquing, response from OpenAI executives).

[160] For an argument from privacy scholars that scraping is "antithetical to privacy" and "violates nearly every key principle embodied in privacy law's frameworks and codes," see Daniel J. Solove & Woodrow Hartzog, *The Great Scrape: The Clash Between Scraping and Privacy*, 113 CAL. L. REV. 1521, 1524 (2025).

[161] *See* Dhruv Mehrotra & Tim Marchman, *Perplexity Is a Bullshit Machine*, WIRED (June 19, 2024, 9:00 AM), https://www.wired.com/story/perplexity-is-a-bullshit-machine/; Robb Knight, *Perplexity AI Is Lying about Their User Agent* (June 14, 2024), https://rknight.me/blog/perplexity-ai-is-lying-about-its-user-agent/.

[162] *See* Dhruv Mehrotra & Andrew Couts, *Amazon Is Investigating Perplexity Over Claims of Scraping Abuse*, WIRED (June 27, 2024, 6:15 PM), https://www.wired.com/story/aws-perplexity-bot-scraping-investigation/.

[163] *Perplexity AI: The Answer Engine with a Lot of Question Marks*, THE VERGE (updated Nov. 18, 2024, 9:00 AM), https://www.theverge.com/24187792/perplexity-ai-news-updates.

not capture the full measure of social cost. As a practical matter, though, any such non-monetary social cost hasn't changed business practices. Instead, the information privacy law status quo permits AI developers to violate privacy norms and still do business as usual.

Accordingly, uneven regulatory costs across copyright law and privacy law create ripe conditions for AI companies to engage in exploitation. The next Part specifies how they might do so before considering why we should care if they do, a topic that Part IV takes up in more detail.

D. The Domain Exploitation Playbook: "Buy" or "Ask"

This Part analyzes how leading AI developers can exploit lower overall legal and social costs in one domain (information privacy law) to diffuse comparatively higher overall legal and social costs in another domain (copyright law). It identifies two leading tactics, "buy" and "ask," that private actors use to leverage doctrinal collapse and assesses how each tactic minimizes overall regulatory costs.[164]

1. Domain Exploitation: The "Buy" Approach

First, a company may buy data through licensing agreements.[165] What typifies a buy is a business-to-business deal that focuses on the relationship between one entity (an AI developer) and another entity (an aggregator of

[164] Others have highlighted how the prospect of legal liability affects the types of data that AI creators rely on. For a prescient early account, see Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, *supra* note 23, at 589 ("The friction caused by copyright law encourages AI creators to use biased, low-friction data (BLFD) for training AI systems."). Professor Levendowski identifies "two ways to acquire . . . [copyrighted works to use as training data for AI systems] without worrying about the threat of copyright infringement: AI creators can build a system to get those works or buy them from someone else." *Id*. at 606 (internal citations omitted). Although the present Article also analyzes two specific tactics that companies use to acquire data to develop AI systems, my account differs in both emphasis and scope. First, whereas Professor Levendowski focuses on AI, biased data, and one domain of law (copyright law), *see id.* at 589, my analysis focuses on the relationship between two domains (information privacy law and copyright law) and the political economy and rule of law consequences of the data acquisition status quo, *see supra* Part III.A-B. Second, the "buy" and "ask" tactics that I identify refer to *all* data acquisition—not only acquisition of copyrighted works. Indeed, one of the reasons that companies might engage in the tactics that I identify is to acquire *other* data, which might include, for instance, data to fine-tune the model after the initial training, thereby obtaining an edge on other AI developers.

[165] *See also* Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, *supra* note 23, at 582, 606 (discussing how AI developers might enter licensing agreements after copyright infringement lawsuits and attempt to "buy . . . [copyrighted works] from someone else").

content).[166] There are at least three potential subtypes of licensing arrangements. The first subtype consists of deals between AI companies and news publishers, where the contributed content is material written by employees of the publisher. Think here of a licensing agreement between, for example, Meta and Reuters,[167] or between OpenAI and Condé Nast.[168] The second subtype, which is rare as of this writing, consists of deals between AI companies and other publishers, where the contributed content comes from non-employee authors. Think here of a licensing agreement between, say, the "Big Five" book publisher Harper Collins and an AI company.[169] The third subtype consists of deals between online platforms and AI companies, where the contributed content is user generated. Think here of a licensing agreement between, for example, a blogging platform like WordPress and Midjourney,[170] or between Shutterstock and OpenAI.[171] In each of these instances, there may be additional negotiations, payments, or other agreements between the company doing the licensing and the individual authors or contributors. For example, Harper Collins indicated that it obtained authors'

[166] Such a buy can take many forms; this tactic includes, for instance, a trade group formed by data licensing companies in the music and image licensing industry, *see* Katie Paul, *AI Dataset Licensing Companies Form Trade Group*, REUTERS (June 26, 2024, 1:05 PM), https://www.reuters.com/technology/artificial-intelligence/ai-dataset-licensing-companies-form-trade-group-2024-06-26/, licensing deals between AI companies like OpenAI and media companies like Vox Media and The Atlantic, *see* Todd Spangler, *OpenAI Inks Licensing Deals to Bring Vox Media, The Atlantic Content to ChatGPT*, VARIETY (May 29, 2024, 8:15 AM), https://variety.com/2024/digital/news/openai-vox-media-atlantic-chatgpt-licensing-deals-1236018547/, licensing deals between technology firms like Google and OpenAI and other internet platforms, like Reddit, *see* Sarah E. Needleman, *Reddit to Give OpenAI Access to Its Data in Licensing Deal*, WALL ST. J. (May 16, 2024, 5:04 PM), https://www.wsj.com/tech/ai/reddit-signs-data-licensing-deal-with-openai-14993757; Anna Tong, Echo Wang & Martin Coulter, *Exclusive: Reddit in AI Content Licensing Deal with Google*, REUTERS (Feb. 21, 2024, 11:10 PM), https://www.reuters.com/technology/reddit-ai-content-licensing-deal-with-google-sources-say-2024-02-22/, alleged further licensing deals between OpenAI and Automattic, the parent company of the blog platforms Tumblr and WordPress, *see* Samantha Cole, *Tumblr and WordPress to Sell Users' Data to Train AI Tools*, 404 MEDIA (Feb. 27, 2024, 1:21 PM), https://www.404media.co/tumblr-and-wordpress-to-sell-users-data-to-train-ai-tools/, and reported licensing deals between technology companies like Apple and news publishers, see Benjamin Mullin & Tripp Mickle, *Apple Explores A.I. Deals with News Publishers*, N.Y. TIMES (Dec. 22, 2023), https://www.nytimes.com/2023/12/22/technology/apple-ai-news-publishers.html.

[167] *See* Mia Sato, *Meta Signs Its First Big AI Deal for News*, THE VERGE (Oct. 24, 2024, 7:46 AM), https://www.theverge.com/2024/10/25/24279259/meta-reuters-ai-chatbot-deal-news-licensing-media.

[168] *See* Kate Knibbs, *Condé Nast Signs Deal with OpenAI*, WIRED (Oct. 20, 2024, 2:00 PM), https://www.wired.com/story/conde-nast-openai-deal/.

[169] *See HarperCollins Inks AI Licensing Deal for Nonfiction Books*, PUBLISHERS WKLY. (Nov. 18, 2024), https://www.publishersweekly.com/pw/newsbrief/index.html?record=5076.

[170] *See* Cole, *Tumblr and WordPress to Sell Users' Data to Train AI Tools*, *supra* note 166.

[171] *See* Press Release, *Shutterstock Partners with OpenAI and Leads the Way to Bring AI-Generated Content to All*, SHUTTERSTOCK (Oct. 25, 2022), https://www.shutterstock.com/press/20435.

opt-in agreement and provided payment for the company's licensing deal,[172] and Shutterstock's announcement of its 2022 licensing deal with OpenAI emphasized the compensation that it would provide to contributors.[173] This user-to-business engagement is an "ask," which I discuss below.[174] The "buy" tactic emphasizes a distinct transaction: The business-to-business deal.

To sharpen the stakes, this Article focuses on licensing arrangements between online platforms and AI companies and emphasizes cases in which the licensed material includes user-generated content that might raise both copyright and privacy concerns for individual users.[175] In such instances, a "buy" reduces overall regulatory costs by taking advantage of limitations and weaknesses in both fields of law.

Begin with the ways in which a "buy" of this sort leverages privacy law. A licensing deal takes advantage of limited legal protections for data once it has been initially disclosed under the notice-and-choice framework. It relies on the underlying terms of service or privacy policy as evidence that the user consented to the deal when they disclosed the data to the business that initially collected it. But the fundamental challenge here, and in any situation in which data disclosed in one setting is licensed for use in AI systems, is that the transaction involves only the businesses—even though the underlying data has changed contexts in ways that quite likely violate the privacy expectations of the individual user who originally disclosed it. American privacy law, as it currently stands, generally doesn't consider publicly disclosed data of the sort that is typically involved in licensing agreements to be covered by the domain of privacy law at all.[176]

OpenAI and Google's respective moves to license data from Reddit illustrate how a buy leverages this underlying weakness in privacy law protections.[177]

---

[172] *See* Andrew Albanese & Jim Milliot, *Agents, Authors Question HarperCollins AI Deal*, PUBLISHERS WKLY. (Nov 19, 2024) https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/96533-agents-authors-question-harpercollins-ai-deal.html (discussing terms of deal).

[173] *See* Press Release, *Shutterstock Partners with OpenAI and Leads the Way to Bring AI-Generated Content to All, supra* note 172.

[174] See *infra* Part II.D.2.

[175] I recognize that different kinds of user-generated material may present distinct privacy and copyright law considerations. For example, if a user uploads a selfie to a social media site, the user has both copyright interests (in the image) and privacy interests (in the biometric data). A user-uploaded photograph of a gorgeous sunset, in contrast, might more naturally present copyright interests, with minimal privacy claims available under contemporary U.S. law. Still, there remains a category of user-contributed, licensed data that presents privacy and copyright concerns with force. Thank you to Kat Geddes for pushing me on this point.

[176] See *supra* note 110 and accompanying text.

[177] On Reddit's agreement with OpenAI, see Sarah E. Needleman, *Reddit to Give OpenAI Access to Its Data in Licensing Deal*, WALL ST. J. (May 16, 2024, 5:04 PM), https://www.wsj.com/tech/ai/reddit-signs-data-licensing-deal-with-openai-14993757. On

When a teenager newly diagnosed with an autoimmune condition posts on Reddit in, say, "r/ChronicIllness," billed as "[a] place of support for those living with, or affected by, chronic illness" that is "[o]pen and welcoming to all,"[178] or a woman struggling with repeat pregnancy loss posts in "r/infertility,"[179] the poster technically chose to disclose this information in public, consistent with the terms of the platform.[180] But it seems unlikely that most of these individuals expected this highly sensitive personal data to be repurposed and used to train an AI model.

To be fair, companies that have entered licensing agreements might respond that the terms of the deal do respect user privacy, within the letter of contemporary privacy law. Indeed, Reddit itself makes such a claim in a June 2025 lawsuit against the AI developer Anthropic, which has not licensed Reddit data and is alleged to have illegally scraped Reddit content to train its AI model.[181] In the complaint, Reddit asserts that it has "established a market for licensing content, through which Reddit imposes meaningful guardrails on the use of such content to protect both Reddit and its users."[182] Assuming that Reddit complies with its own terms of service (ToS) and that there is user consent, Reddit does satisfy American privacy law's formal demands. But guardrails or not, a "buy" entails a business-to-business deal that elides further analysis of user privacy interests. That's the point. The structure of privacy law on the books permits such a licensing deal, without imposing much—if any—legal cost in the privacy law domain. Current understandings of "public" permit analysis of privacy interests in the data to collapse into copyright law arguments about licensing of the data.

In addition, even if a "buy" involves a media entity or online publisher and not a platform company, such that there are arguably no individual privacy interests in the data, there may still be incentives for well-resourced companies

---

Reddit's agreement with Google, see Annelise Gilbert, *Google-Reddit AI Deal Heralds New Era in Social Media Licensing*, BLOOMBERG L. (March 7, 2024, 2:06 AM), https://news.bloomberglaw.com/ip-law/google-reddit-ai-deal-just-the-start-for-social-media-licensing.

[178] *r/ChronicIllness*, REDDIT, https://www.reddit.com/r/ChronicIllness/.

[179] *r/infertility*, REDDIT, https://www.reddit.com/r/infertility/.

[180] *See Reddit Privacy Policy,* REDDIT (May 29, 2025), https://www.reddit.com/policies/privacy-policy ("Much of the information on the Services is public and accessible to everyone, even without an account. . . . You should take the public nature of the Services into consideration before posting. By using the Services, you are directing us to share this information publicly and freely.").

[181] *See* Reddit, Inc. v. Anthropic PBC, No. 3:25-cv-05643 (N.D. Cal. Jul. 3, 2025)*; see also* Mike Isaac, *Reddit Sues Anthropic, Accusing It of Illegally Using Data from Its Site*, N.Y. TIMES (June 4, 2025), https://www.nytimes.com/2025/06/04/technology/reddit-anthropic-lawsuit-data.html (discussing lawsuit).

[182] Complaint ¶ 10, Reddit, Inc. v. Anthropic PBC, No. CGC-25-625892 (N.D. Cal. June 4, 2025).

to normalize licensing as a socially acceptable move. Part IV.B returns to the broader political economy implications of this tactic.

The current structure of copyright law also makes the "buy" tactic attractive for at least two reasons. First, a "buy" can also avoid legal costs because a company that lawfully acquires data in a licensing deal will not expose itself to claims of unlawful copyright infringement. Indeed, a failed attempt to reach a licensing agreement can result in a lawsuit, as was the case in *New York Times v. OpenAI*.[183] Second, a "buy" can reduce social costs. An entity that buys data can boast that it respects artists and wants to ensure that there is fair compensation for artists' labor, diffusing the force of equity-based copyright law arguments. Even if that money flows to the content aggregator (an entity) and not directly to individual content creators, the AI developer can justify its actions by contending that it is not just legally compliant—it is going above and beyond and doing the right thing.[184] The company can thereby retain social license. An AI developer that buys data can, moreover, potentially assert that it is respecting both copyright and privacy interests by licensing rather than scraping data,[185] further supporting the contention that it is an accountable actor worthy of public trust. The regulatory status quo thus makes the "buy" tactic tempting.

---

[183] N.Y. Times v. Microsoft Corp., 757 F.Supp.3d 594 (S.D.N.Y. 2023); *see* Benjamin Mullin, *Inside the News Industry's Uneasy Negotiations with OpenAI*, N.Y. TIMES (Dec. 29, 2023), https://www.nytimes.com/2023/12/29/business/media/media-openai-chatgpt.html.

[184] *See, e.g.*, Nico Grant & Cade Metz, *The Push to Develop Generative A.I. Without All the Lawsuits*, N.Y. TIMES (July 19, 2024), https://www.nytimes.com/2024/07/19/technology/generative-ai-getty-shutterstock.html (discussing Getty Images' partnership with Picsart, which relies on licensing to "build[] an A.I. image model with stock photos from Getty's repository" and avoid legal controversies). *Cf. Stack Overflow and Google Cloud Announce Strategic Partnership to Bring Generative AI to Millions of Developers*, PR NEWSWIRE (Feb. 29, 2024), https://www.prnewswire.com/news-releases/stack-overflow-and-google-cloud-announce-strategic-partnership-to-bring-generative-ai-to-millions-of-developers-302075701.html (press release announcing "strategic partnership" between Google and Stack Overflow and quoting Stack Overflow CEO: "This landmark, multi-dimensional AI-focused partnership . . . underscores our joint commitment to unleash developer creativity, unlock productivity without sacrificing accuracy, and deliver on socially responsible AI"); Benjamin Mullin & Tripp Mickle, *Apple Explores A.I. Deals with News Publishers*, N.Y. TIMES (Dec. 22, 2023), https://www.nytimes.com/2023/12/22/technology/apple-ai-news-publishers.html ("Two people familiar with [Apple's licensing] discussions struck a positive note on the long-term prospects of a deal, contrasting Apple's approach of asking for permission with . . . other artificial intelligence-enabled companies, which have been accused of seeking licensing deals with news organizations after they had already used their content to train generative models.").

[185] *See, e.g.*, Mullin & Mickle, *Apple Explores A.I. Deals with News Publishers*, *supra* note 184 (discussing Apple's licensing negotiations with major news organizations and reporting that "Apple has been reluctant to take information from the internet, partly because of its commitment to privacy").

2. Domain Exploitation: The "Ask" Approach

Alternatively or in addition, an AI developer might pursue a distinct tactic to acquire data: Engage directly with an individual and "ask" that person to consent to the use of their data for AI training through privacy policies, terms of service, or both.[186] In the privacy law domain, an "ask" occurs when AI companies exploit the lower regulatory costs created by American privacy law's notice-and-choice approach.[187] The privacy policies for many tech companies now disclose that user data may be used to develop AI products, serving as the notice of the dominant notice-and-choice framework.[188] As for the choice,

---

[186] By calling this an "ask," I do not mean to suggest that the user feels they are being given a free choice or that a user's silence or choice not to opt out amounts to informed, express consent.

[187] Others have previously recognized the relationship between tech companies' business models and data acquisition, including in the AI context. For instance, an "ask" can at times overlap with Professor Levendowski's "build-it" approach, which entails "amassing training data from users in exchange for a service those users want." Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, *supra* note 23, at 606 (citing Katherine J. Strandburg, *Monitoring, Datafication, and Consent: Legal Approaches to Privacy in the Big Data Context*, *in* PRIVACY, BIG DATA, AND THE PUBLIC GOOD: FRAMEWORKS FOR ENGAGEMENT 10 (Julia Lane et al. eds., 2014), http://wpressutexas.net/cs378h/images/b/b3/LaneEtAlPrivacyBigDataAndThePublicGood.pdf). The "ask" tactic that I identify in this Article refers specifically to data acquired through direct interactions with the user, in which the company seeks consent for that collection through either an "opt in" or "opt out" arrangement, as opposed to more generally referencing private entities' business models. *See id.* at 606 n.128 (quoting Evgeny Morozov (@evgenymorozov), X (Jan. 20, 2018, 12:40 AM), https://x.com/evgenymorozov/status/954634817198546945? (referencing Baidu chief scientist Andrew Ng, who "restated the build-it model: '[a]t large [tech] companies, we often launch products not for the revenue, but for the data . . . and we monetize the data through a different product'")).

[188] *See, e.g.*, *Privacy Policy*, GOOGLE (effective Dec. 11, 2025), https://policies.google.com/privacy (AI-related information under "Research and Development" portion of "Compliance & Cooperation with Regulators" section); *Privacy Policy*, LINKEDIN (effective Nov. 3, 2025), https://www.linkedin.com/legal/privacy-policy (AI-related information under Section 2, How We Use Your Data); *see also LinkedIn and Generative AI (GAI) FAQs*, LINKEDIN, https://www.linkedin.com/help/linkedin/answer/a5538339?hcppcid=search (last visited Dec. 26, 2025) (describing how LinkedIn may process user data to train and improve generative AI models); *Privacy Policy*, META, https://mbasic.facebook.com/privacy/policy/printable/ (effective Dec. 16, 2025) (AI-related information provided in multiple, often cross-referenced locations, including "What information do we collect?" under "Your activity and information you provide" (providing "learn more" link that, if clicked, cross-references "Information you or others exchange with AI at Meta" section that lists examples, including prompts, AI responses, actions the user asks AI at Meta to take, and user feedback); "How do we use your information?," "Information you or others exchange with AI at Meta" and "How do we share information with third parties?" under "Third Parties," "AI Integrations" section); *see also Privacy notice for United States residents*, META, https://mbasic.facebook.com/privacy/policy/printable/ (effective Dec. 16, 2025) (AI-related

policies vary, but they are generally limited.[189] Notably, for LLM-based chatbots, a 2025 study of six American developers found that "all use user chats (inputs) with their chatbots by default to train their LLMs."[190] When it comes to the broader category of companies developing AI, options remain limited. For

information available under "Developing and improving AI at Meta"); *How Meta uses information for generative AI models and features*, AI at Meta, https://www.facebook.com/privacy/genai (last visited Dec. 26, 2025) (describing how Meta uses "information that is publicly available online and licensed information" as well as "information shared on Meta Products" for AI training); *X Privacy Policy*, X, https://x.com/en/privacy (effective Jan. 15, 2026) (last visited Dec. 26, 2025) (AI-related information under "2. How We Use Information," section 2.1, "Operate, improve, and personalize our services"). The same is true for companies focused on AI development. *See, e.g.*, *Privacy Policy*, ANTHROPIC, https://www.anthropic.com/legal/privacy (effective Oct. 8, 2025) (discussion of collection and processing of personal data to train model under multiple sections, including "Collection of Personal Data," subsection on "Personal data we collect or receive to train our models" and "Legitimate Bases for Processing"); *Privacy Policy*, OPENAI, *supra* note 117. For reporting on this development, see Eli Tan, *When the Terms of Service Change to Make Way for A.I. Training*, N.Y. TIMES (June 26, 2024), https://www.nytimes.com/2024/06/26/technology/terms-service-ai-training.html. For discussion of LinkedIn's policy changes in September 2024, see Chris Velazco, *LinkedIn is Training AI on You — Unless You Opt out with This Setting*, WASH. POST (Sep. 23, 2024), https://www.washingtonpost.com/technology/2024/09/23/linkedin-training-ai-setting-opt-out/. Similar dynamics are present in other companies that one might not think of as "AI companies" or "big tech" firms, such as the video conferencing service Zoom, the question-and-answer site Quora, and the messaging service Slack. *See* Matt Burgess & Reece Rogers, *How to Stop Your Data from Being Used to Train AI*, WIRED (Oct. 12, 2024, 9:30 AM), https://www.wired.com/story/how-to-stop-your-data-from-being-used-to-train-ai/ (reporting on Quora and Slack's AI policies); Ivan Mehta & Ingrid Lunden, *Slack Under Attack over Sneaky AI Training Policy*, TECH CRUNCH (May 17, 2024, 8: AM PDT), https://techcrunch.com/2024/05/17/slack-under-attack-over-sneaky-ai-training-policy/ (reporting on Slack's policy); Melissa Goldin, *Zoom Says It Isn't Training AI on Calls Without Consent. But Other Data Is Fair Game*, AP NEWS (Aug. 9, 2023, 5:55 AM), https://apnews.com/article/fact-check-zoom-ai-privacy-terms-of-service-06ff47e47439c2173390a4ca1389f652 (reporting on Zoom's policy).

[189] This reporting reflects my best efforts to trace corporate policy as of this writing; however, it is quite challenging, even for experts, to trace company policy on the processing of user data to train generative AI models. *See* Jennifer King , Kevin Klyman, Emily Capstick, Tiffany Saade, & Victoria Hsieh, *User Privacy and Large Language Models: An Analysis of Frontier Developers' Privacy Policies*, *in* Proceedings of the Eighth AAAI/ACM Conference on AI, Ethics, and Society (AIES 2025), at 1466 (studying six U.S.-based chatbot developers and finding that they all "rely upon a web of documents in addition to their primary privacy policies to govern their use of users' chat data"). This lack of clarity on what companies are doing heightens the rule of law concerns discussed *infra* Part IV.B.

[190] King et al, *supra* note 189, at 1465–66.

example, Anthropic, Google, and OpenAI provide a partial opt-out right for users;[191] Meta does not.[192]

Moreover, similarly limited user options also appear in the copyright law domain, where a company's "ask" can serve as a form of license for data acquisition.[193] Such a license can either be implied through a privacy policy or made explicit through separate ToS. Google's ToS, which provides that the company can use content contributed by its users for "developing new technologies and services," is illustrative of an explicit "ask" of this sort.[194]

---

[191] The opt-out options vary. In the case of OpenAI, individual users of the freely available model are now given the choice to opt-out, albeit with language warning that they won't be helping to "improve the model for everyone." *See Data Controls FAQ*, OPENAI, https://help.openai.com/en/articles/7730893-data-controls-faq (last updated Dec. 26, 2025) (describing steps to "stop my chats from training ChatGPT? (ie. 'Improve the model for everyone')); *see also Consumer Privacy at OpenAI*, OPENAI (June 12, 2024), https://openai.com/consumer-privacy/ (describing "data controls:" "Chat data can help us improve model quality, like how to give clearer answers to questions people ask every day—but only if you want that. You can turn this setting off at any time . . ."). Anthropic similarly permits individual users of the Claude Free, Max, and Pro plans to opt out by toggling a switch labelled "You can help improve Claude" to "off." *See Updates to Consumer Terms and Privacy Policy*, ANTHROPIC https://www.anthropic.com/news/updates-to-our-consumer-terms (Aug. 28, 2025) (describing updates to consumer terms to "giv[e] users the choice to allow their data to be used to improve Claude"). In the case of Google, the picture is mixed: If your data is part of "publicly available information," then it is used to "help train Google's AI models and build products and features . . .". *Google Privacy Policy & Terms*, GOOGLE, https://policies.google.com/privacy (effective Dec. 11, 2025). However, Google's "personal AI assistant", Gemini Apps, allows users to opt out of the "keep activity" setting, which is turned on by default for users outside of the European Economic Area, Switzerland, and the UK. *See Gemini Apps Privacy Hub*, GOOGLE (Dec. 23, 2025), https://support.google.com/gemini/answer/13594961#gemini_app ("How can I control whether Gemini Apps use my data to personalize responses?"; Gemini Apps Activity, GOOGLE, https://myactivity.google.com/product/gemini (last visited Dec. 26, 2025) (providing click-through menu to turn "off" setting under "Keep Activity" heading). Unless this setting is turned off by the user, Gemini Apps "uses your activity to provide, develop, and improve its services (including training generative AI models), as well as to protect Google, its users, and the public with the help of human reviewers." *Gemini Apps Privacy Hub*, *supra.* One further caveat is that "audio and Gemini Live recordings aren't used to improve Google services by default," meaning that the user must opt-in to these uses. *Id.* Google's Gemini Apps Help page also notes that, "[d]epending on your settings and region, Google also uses your activity to personalize your experience." *Id*. Changes made to Gemini's privacy settings in late 2025 are the subject of a pending case, *Thele*, discussed *supra* note 142.

[192] In the case of Meta, there is no way for most American users to use Meta's suite of services, including Facebook and Instagram, without consenting to the company's processing of their data, including for Meta's generative AI products. *See Privacy Notice for United States Residents*, META, *supra* note 188.

[193] Thank you to BJ Ard and Bob Brauneis for especially helpful conversations about this portion of the argument. On generative AI, terms of service, and copyright licenses, see Kim, *supra* note 54, at 594. On AI companies' attempts to use ToS to restrict certain uses of their AI models or outputs, see Henderson & Lemley, *supra* note 54, at 1340-45.

[194] *Google Terms of Service: Permission to Use Your Content*, GOOGLE (May 22, 2024) (United States version), https://policies.google.com/terms?hl=en-US#toc-permission.

Accordingly, thin, contract-based consent is used both in privacy law and, in at least some cases, to obtain copyright permissions for material created or shared by users.

An "ask" to acquire user data allows a company both to limit future exposure to copyright liability and to mitigate copyright-adjacent social costs, while still obtaining the data that the company needs to develop AI systems. Because high costs in copyright law come from allegations of direct infringement when companies scrape "publicly available" data to train their models,[195] "asking" a user for their data diffuses this legal cost. And because the data scraped to train AI models is also at the center of public controversies about labor and equity,[196] "asking" a user for their data diffuses this social cost.

To be sure, an "ask" of this sort will not always eliminate all social contestation. Picture a situation in which a group of content creators collectively protests a platform's policy on the use of their data as inputs to train its AI model. If the platform prizes its relationship with the community of users, and if enough individuals can band together, then they may be able to increase social costs for the platform.[197] But especially if these content creators lack clout, money, or

[195] Available evidence suggests that this training data was obtained by scraping "publicly available" material on the internet. *See supra* Part II.C (discussing scraping).

[196] *See id.*

[197] For two examples in which non-famous content creators were able to generate concentrated social pressures, consider Adobe and SoundCloud. In 2024, Adobe came under fire for changes to its terms of service that seemed to authorize AI training on user-contributed content. *See* Tiffany Ng, *Adobe Says It Won't Train AI Using Artists' Work. Creatives Aren't Convinced.*, WIRED (Jun. 19, 2024, 1:59 PM), https://www.wired.com/story/adobe-says-it-wont-train-ai-using-artists-work-creatives-arent-convinced/. For a sense of long-simmering customer anger with Adobe's use of generative AI, see *Adobe Gives Middle Finger to All Their Human Customers with Text to Image Software*, ADOBE COMMUNITY, https://community.adobe.com/t5/adobe-firefly-discussions/adobe-gives-middle-finger-to-all-their-human-customers-with-text-to-image-software/m-p/14175535 (2023 discussion forum protesting Adobe's use of generative AI for text-to-image content creation). The company subsequently clarified that it only uses AI training for "[c]ontent you choose to submit to the Adobe Stock marketplace," which "is governed by the separate Adobe Stock Contributor Agreement." *See* Adobe General Terms of Use, ADOBE (effective Oct. 3, 2025), https://www.adobe.com/legal/terms.html. In addition, early summer 2025, in the wake of public outcry about the potential use of user-contributed artistic content to train generative AI models, the music-sharing site SoundCloud changed its terms of service and "ma[de] a formal commitment that any use of AI on SoundCloud will be based on consent, transparency, and artist control." Press Release, A Letter from our CEO: Clarifying our Terms of Use, SOUNDCLOUD (May 14, 2025), https://press.soundcloud.com/249951-a-letter-from-our-ceo-clarifying-our-terms-of-use. The CEO asserted that the change was to clarify the company's longstanding practice: "SoundCloud has never used artist content to train AI models. Not for music creation. Not for large language models. Not for anything that tries to mimic or replace your work. Period. We don't build generative AI tools, and we don't allow third parties to scrape or use artist content from SoundCloud to train them either." *Id.*

other forms of influence,[198] then the company is more likely to be able to retain social license, even without changing its business practices. This combination of factors makes an "ask" an appealing way to reduce copyright-related legal costs and simultaneously allow the company to lessen the risk of copyright-related social costs, without incurring major legal or social costs in privacy law.

In theory, the prospect of legal liability for this conduct or an even stronger social reaction could change the regulatory calculus in privacy law. But lawsuits targeting this sort of conduct have not gotten far,[199] and contemporary companies seem to retain their social license even when they rely on an "ask" to acquire data for AI systems.[200] If the legal framework were different, then the privacy law costs might be higher. But they aren't. Compare the data protection regulatory regime in the EU, where AI developers do not rely on this form of data collection to train AI models and must ensure, among other requirements, that there is a lawful basis for each distinct phase of data processing (including collection).[201] This legal requirement, imposed by the General Data Protection

---

[198] *See generally* Marc Galanter, *Why the "Haves" Come Out Ahead: Speculations on the Limits of Legal Change*, 9 L. & SOC. REV. 95 (1974) (distinguishing "haves" from "have-nots" based on power, wealth, and status).

[199] For example, the plaintiffs in one proposed class action lawsuit against LinkedIn for its privacy policy changes filed their complaint on January 21, 2025, and then filed a notice to dismiss the claims just nine days later. *See* Complaint, Torre v. LinkedIn Corp, No. 5:25-00709, Doc. No. 1 (N.D. Cal. Jan. 21, 2025); Notice of Dismissal Without Prejudice, De La Torre v. LinkedIn Corp, No. 5:25-00709, Doc. No. 7 (N.D. Cal. Jan. 30, 2025).

[200] See discussion *supra* Part III.C.

[201] Under the GDPR, a company must, among other requirements, have a specific lawful basis for data collection. *See* Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC, 2016 O.J. (L 119) (repealing General Data Protection Regulation, art. 4(2) (defining data processing to include collection), and art. 6(1) ("Processing shall be lawful only if and to the extent that at least one of the following [conditions] applies."). To date, AI developers have generally relied on "performance of a contract" and "legitimate interests" as the lawful bases for generative AI model development and deployment. *See, e.g.*, *Europe Privacy Policy*, OPENAI (Nov. 4, 2024), https://openai.com/policies/eu-privacy-policy/ (describing OpenAI's "legitimate interest" as a "legal basis" for processing user data to improve services, conduct research, prevent fraud or misuse, and protect the privacy and safety of users); *Gemini Apps Privacy Hub*, *supra* note 191 (indicating legal bases for Gemini, "your personal AI assistant from Google," to process data and listing "performance of a contract" and "Google and third parties' legitimate interests with appropriate safeguards to protect your privacy," as well as "legal obligations," in certain cases, and "your consent," in the case of "certain features"). A December 2024 opinion by the European Data Protection Board, which interprets the GDPR, "provides general considerations . . . to take into account when assessing whether controllers can rely on legitimate interest as an appropriate legal basis for processing conducted in the context of the development and the deployment of AI models." *See* Eur. Data Prot. Bd., *Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models*, at 2 (Dec. 17, 2024),

Regulation,[202] raises the regulatory costs of data acquisition. The U.S. regulatory regime, however, does not entail the same legal costs. And to date, there had not been enough of social pushback against AI developers' privacy practices to move the needle.[203] Because privacy law's overall social and legal costs are lower than the overall legal and social costs in copyright law, an "ask" to acquire data from users remains an appealing tactic for AI developers.

* * *

The "ask" and "buy" tactics demonstrate how doctrinal collapse enables exploitation. Collapse creates structural conditions for firms to capitalize on blurring legal boundaries.[204] One possible response to a "buy" or an "ask" is a shrug. In less glib terms, one might respond that exploitation of doctrinal lines makes it easier to amass the data required to develop AI tools. One might further contend that these tactics are good, because they route around formal doctrinal lines and rigid doctrinal logics that might otherwise hamper salutary technological innovation. For example, machine-learning tools can "expose hidden discrimination in social systems."[205] Take healthcare: targeted development of advanced digital technologies might correct historic patterns of racial inequity.[206] Emerging data-driven technologies might also permit ground-breaking medical insights. Consider AlphaFold, an AI system developed by Google DeepMind that can predict amino acid protein structures and promises to advance biomedical research.[207] This work solved a 50-year old problem and is so revolutionary that it was awarded a 2024 Nobel Prize in Chemistry.[208] Furthermore, the potential benefits of AlphaFold are not locked up to benefit only select private actors; rather, Google has partnered with EMBL's European Bioinformatics Institute to create AlphaFold DB, which "provides open access to over 200 million protein structure predictions to accelerate scientific

https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf at 2.

[202] *See* Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC, *supra* note 201, at art. 4(2) & art. 6(1).

[203] *See supra* Part II.C.

[204] See *supra* Part I.A.

[205] *See, e.g.*, Solow-Niederman, *Information Privacy and the Inference Economy*, *supra* note 71, at 402.

[206] Solow-Niederman, *Information Privacy and the Inference Economy*, *supra* note 71, at 402-03.

[207] *See* Josh Abramson et al., *Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3*, 640 NATURE 493 (2024).

[208] *See* Press Release, Royal Swedish Acad. Sci., The Nobel Prize in Chemistry 2024 (Oct. 9, 2024), https://www.nobelprize.org/uploads/2024/10/press-chemistryprize2024-3.pdf.

research."[209] Developments like these suggest the potential good that AI-driven insights can produce. If collapse enables exploitation that, in turn, produces this kind of beneficial technological innovation, then some might say that there is no problem at all.

From this perspective, focusing on whether doctrinal collapse and associated exploitation are or are not problematic, full stop, trains attention on the wrong normative question. The issue is not the tools, but rather how they are used. The right question to ask is how technological innovations that rely on data are applied, in social context, and who is (not) served by these interventions.[210]

At least at present, inter-regime doctrinal collapse enables forms of exploitation that are not evenly distributed. As the next Part argues, benefits flow disproportionately to well-resourced private firms, whereas the costs are disproportionately borne by the public (both in the sense of costs for individuals and costs for the broader legal system). Even for those who disagree, framing the consequences of collapse in this way underscores the broader governance challenge, which Part V takes up: Whether it is possible to structure legal institutions to permit desirable forms of regulatory and technological innovation, while amply constraining the arbitrary exercise of private power.

III. The Consequences of Doctrinal Collapse

This Part moves from what doctrinal collapse is, to why it matters. It contends that, at least when it comes to data acquisition, doctrinal collapse's potential upsides do not outweigh the individual and systemic costs. Part IV.A focuses on the political economy of AI development, asserting that collapse enables patterns of exploitation that are apt to prioritize larger players, at the expense of the (non-famous) user. Part IV.B turns to collapse's jurisprudential toll, underscoring the harms of legal incoherence and the cost to democratic legitimacy.

A. The Political Economy of Collapse: Who Wins and Who Loses

[209] ALPHAFOLD PROTEIN STRUCTURE DATABASE, https://alphafold.ebi.ac.uk/ (last visited May 24, 2025).

[210] Solow-Niederman, *Information Privacy and the Inference Economy*, *supra* note 71, at 403 ("Asking whether a tool helps or harms is the wrong question. The better set of questions is: who does the tool purport to help, with what costs, and how are the costs and benefits distributed?"); *see also* Jack M. Balkin, *The Path of Robotics Law*, 6 CALIF. L. REV. CIR. 45, 48 (2015) ("[T]he most important lesson of cyberlaw for robotics is the need to attend to the relationships between affordance and imagination, between tools and relations of power, between technological substrate and social use.").

This Part confronts the distributional consequences of collapse, contending that the collapse of IP and privacy law in AI development favors the "haves."[211] On balance, it's easier for larger, more well-established players to deploy each of the exploitation tactics identified in Part III.[212]

1. Who Can "Buy"

The "buy" tactic isn't equally available to all actors, and it disproportionately favors the interests of big business. First, not every entity is well-positioned to license data. The "buy" strategy is likely to favor larger players with adequate market power to negotiate favorable licensing terms. A start-up may lack the resources to negotiate licensing deals with a major internet platform or media company. But OpenAI and other similarly positioned entities can.[213] For those who can afford to execute it, a "buy" tactic is attractive because it allows the company both to avoid legal costs, in the form of copyright litigation, and to diffuse social costs, in the form of equitable concerns about copyright law and IP rights, more generally.

Second, not everyone feels the effects of a "buy" the same way. Again, this point is sharpest in the context of licensing arrangements that involve platforms with user-generated content (as compared to licensing arrangements that involve media entities).[214] The typical platform user is not engaged in the business-to-business "buy" and can do little to contest a licensing scheme that involves their work, to which they are bound by the platform's terms of service. In some cases, enough public controversy might lead to changes in the ToS.[215] But there is no guarantee.[216] By way of further illustration, if I am an amateur musician who records songs for my family, I might protest if I upload a song to a music platform and later learn that, consistent with its ToS, the platform has used my song to train its generative AI model. But even without bringing industry trade

---

[211] *See* Galanter, *supra* note 198.

[212] *Cf.* Jack M. Balkin, *Room for Maneuver: Julie Cohen's Theory of Freedom in the Information State*, 6 JERUSALEM REV. L. STUD. 79, 81, 85 (2012) (reviewing JULIE E. COHEN, CONFIGURING THE NETWORKED SELF (2012), and predicting that private actors would exploit "semantic discontinuity," meaning "gaps and inconsistencies within systems of meaning," for their own benefit).

[213] See discussion *supra* Part III.A (describing companies that have relied on a buy to acquire data).

[214] *See supra* note 175 and accompanying text.

[215] *See id.*; *see also* Jess Weatherbed, *Adobe's New Terms of Service Aren't the Problem—It's Trust*, THE VERGE (June 7, 2024, 12:37 PM), https://www.theverge.com/2024/6/7/24173838/adobe-tos-update-firefly-generative-ai-trust (describing public controversy around Adobe Firefly and Adobe's announcements on the use of content to train AI models).

[216] *See supra* text accompanying notes 197–198.

groups or corporate rightsholders into the picture, I don't have a major financial stake, nor a national platform, in the same way that a famous musician does. There's just not much noise that I can make, on my own, and I don't have many other options to share my music. Cumulatively, these dynamics suggest that the "buy" strategy is likely to lead the rich to get richer—while remaining unattainable to other companies seeking to develop AI and disproportionately affecting users without connections.

2. Who Can "Ask"

An "ask" is also only open to a select group of private actors. Because AI is, at least for now, still a big data game,[217] an "ask" favors companies with, one, a large user base and, two, a product or service that easily facilitates informational capitalism-style collection of data about users.

To illustrate the point that only some companies are well-positioned to "ask," compare two companies at the forefront of AI development: Meta and Nvidia. Meta is a social media company that has made substantial investments in AI systems.[218] Nvidia is a computing hardware company that is one of the world's most valuable companies; in July 2025, it became the first publicly traded company with a $4 trillion market valuation.[219]

The two companies have very different missions, as well as very different business models. Meta, which owns Facebook and Instagram, aims to "build[] technology that connects you to people, interests and experiences that matter to you."[220] Its core offering is a service: The social media platform has over 3 billion users as of early 2025.[221] Nvidia, in contrast, offers a product: Its importance for AI development comes from its dominance of the market for AI hardware.[222] The

---

[217] See sources cited *supra* note 103.

[218] *See* Jaspreet Singh, *Meta to Spend Up to $65 Billion This Year to Power AI Goals, Zuckerberg Says*, REUTERS (Jan. 24, 2025), https://www.reuters.com/technology/meta-invest-up-65-bln-capital-expenditure-this-year-2025-01-24/; *see also* Mike Isaac & Cade Merz, *In Pursuit of Godlike Technology, Mark Zuckerberg Amps Up the A.I. Race*, N.Y. TIMES (June 27, 2025), https://www.nytimes.com/2025/06/27/technology/mark-zuckerberg-meta-ai.html.

[219] Michael Liedtke & AP, *AI Kingpin Nvidia Crowned as First Public Company with a $4 Trillion Valuation*, ASSOC. PRESS (July 9, 2025, 1:44 PM), https://apnews.com/article/nvidia-4-trillion-chipmaker-7947e86a7ee9a994b9f16c3c0779b74f.

[220] About, META, https://www.meta.com/about/company-info/ (last visited Oct. 2, 2025).

[221] Stacy Jo Dixon, *Most Used Social Networks 2025, By Number of Users*, STATISTA (Mar. 26, 2025), https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

[222] *See* Liedtke & AP, *supra* note 219; *see also* Our Story 2025, NVIDIA, https://images.nvidia.com/aem-dam/Solutions/homepage/pdf/NVIDIA-Story.pdf ("NVIDIA is the world leader in accelerated computing. We are fundamentally changing how computing

company produces a special type of chip that is one of the most important components of today's AI systems, and its market dominance comes from the sale of these chips to other businesses.[223] Market differences aside, Meta and Nvidia have a key commonality: Each is developing generative AI models, and each is named in (separate) lawsuits alleging copyright infringement in the development of their respective generative AI models.[224]

Suppose that, given the substantial legal and social costs of these copyright lawsuits, Meta and Nvidia each seek alternate data sources to refine and develop AI models. Although both of these companies are powerful "haves," Meta is far better positioned to deploy the "ask" tactic and rely on data acquired through privacy policies and boilerplate terms of service. Meta's superior positioning reflects its business model, which centers on a direct relationship with individual users of its social media platform. Nvidia's privacy policy can and does state that it relies on "public images and other data through sensors on NVIDIA or partner-identified vehicles and in clearly disclosed public spaces–as well as datasets collected by others–to improve the safety and reliability of our [autonomous vehicle] and AI models."[225] Perhaps this specialized dataset is sufficient to allow Nvidia to refine and fine-tune its models. But leveraging this "ask" as a less risky way to obtain the data needed for a competitive edge seems less likely to succeed for Nvidia, for the simple reason that it is a computing hardware company. As a matter of common sense, such an entity has far fewer users and far less data about those users to collect, as compared to a major social media platform.

Put simply, a "big tech" company like Meta is better positioned to "ask" for user data. Meta can not only adjust the privacy policies for entire social media

works and what computers can do. The next industrial revolution has begun.") (last visited Oct. 2, 2025).

[223] Liedtke & AP, *supra* note 219. This compute-driven business model is why some reporting expressed concern about Nvidia's September 2025 partnership with OpenAI. The worry is that Nvidia's $100 billion investment in OpenAI amounts to a circular deal that will help Nvidia's bottom line by increasing demand for its chips: Through this "strategic partnership," OpenAI will "'build and deploy at least 10 gigawatts of AI datacenters with NVIDIA systems," with Nvidia's investment provided "progressively as each gigawatt is deployed." Press Release, OpenAI and NVIDIA announce strategic partnership to deploy 10 gigawatts of NVIDIA systems, OPENAI (Sep. 20, 2025), https://openai.com/index/openai-nvidia-systems-partnership/. *See also* Berber Jin & Robbie Whelan, *Nvidia to Invest Up to $100 Billion in OpenAI*, WALL ST. J. (Sep. 22, 2025, 2:13 PM), https://www.wsj.com/tech/nvidia-openai-100-billion-deal-data-centers-d2f85cae (describing "circular deal" between Nvidia and OpenAI).

[224] *See, e.g.*, Order Denying the Plaintiffs' Motion for Partial Summary Judgment and Granting Meta's Cross-Motion for Partial Summary Judgment, Meta v. Kadrey, No. 23-cv-03417 (N.D. Cal. June 25, 2025) (lawsuit against Meta); Complaint, Nazemian v. Nvidia Corp., No. 3:24-cv-01454 (N.D. Cal. Mar. 8, 2024) (lawsuit against Nvidia).

[225] Privacy Policy, NVIDIA, https://www.nvidia.com/en-us/about-nvidia/privacy-policy/ (effective Sept. 22, 2025) (stated under "For Autonomous Vehicle (AV) and Artificial Intelligence (AI) Research and Development" subheading).

platforms, like Facebook,[226] but also for a slew of related products. For example, in April 2025, the company sent an email to Ray-Ban Meta smart glasses owners announcing that users could no longer opt out of storing voice recordings in the cloud; instead, users must manually delete recordings.[227] Meta's "Voice Controls Privacy Notice," updated at the same time, provides that "[v]oice transcripts and stored audio recordings are otherwise stored for up to one year to help improve Meta's products."[228]

Meta's actions with Ray-Ban represent a canonical "ask." Meta touts its reliance on a "team of vetted and trained personnel who assist in reviewing stored audio recordings of your voice interactions, for purposes of improving Meta's products" and its "compl[iance] with strict privacy and security requirements" in handling user information.[229] The user is left with only the existing notice-and-choice framework of privacy law as the means to check this data acquisition. In so doing, Meta operates within the letter of privacy law, avoids legal and social costs in IP law, and still acquires the data it needs to produce AI models.

Moreover, there are signs that other AI leaders are trying to leverage the "ask" tactic by expanding in ways that would allow them to access user data. Take OpenAI. In July 2025, Reuters reported that OpenAI would launch a web browser.[230] If OpenAI becomes a web browser, then it has a new, expanded set of relationships with users, and it can acquire a whole new set of user data.[231] Suddenly, it can not only "ask" users of its AI tools to opt into data sharing "to improve the model for everyone," but also leverage its browser's ToS and privacy policy to acquire additional data from every person who uses the OpenAI browser for an internet search.[232] What's more, if OpenAI continues to expand into agentic AI, such as tools that navigate the internet on the user's behalf to

---

[226] *See* discussion *supra* Part II.D.

[227] *See* Chris Welch, *Meta Tightens Privacy Policy Around Ray-Ban Glasses to Boost AI Training*, THE VERGE (Apr. 30, 2025), https://www.theverge.com/news/658602/meta-ray-ban-privacy-policy-ai-training-voice-recordings/.

[228] AI Glasses Voice Privacy Notice, META (effective July 22, 2025), https://www.meta.com/legal/ai-glasses/voice-controls-privacy-notice/.

[229] *Id.*

[230] *See* Kenrick Cai, Krystal Hu, & Anna Tong, *OpenAI to Release Web Browser in Challenge to Google's Chrome*, REUTERS (July 9, 2025, 8:16 PM), https://www.reuters.com/business/media-telecom/openai-release-web-browser-challenge-google-chrome-2025-07-09/.

[231] Because such a move would position OpenAI to challenge the Google Chrome browser, there are obvious market incentives to make this move, not to mention potential competition law implications. I reserve these points to focus on data acquisition tactics.

[232] Cai et al., *supra* note 230 (". . . [The web browser] will give OpenAI more direct access to a cornerstone of Google's success: user data.").

accomplish a task,[233] and the company relies on its web browser as a gateway for its "agents," then it has gained access to yet more user data. All that OpenAI has to do is "ask." And the more successful the company is, the more it creates an ecosystem in which users continue to engage in ways that allow the company to acquire more and more data from them. This approach is not limited to OpenAI, either; this specific form of "ask" may be a pathway for other, sufficiently well-resourced AI companies, like Perplexity,[234] which are also moving into the browser and agentic AI space.[235]

Because only certain companies have relationships with users and business models that rely on user data in ways that naturally enable an ask, certain AI companies—those who are already big tech, and especially those who were big tech before AI or who have other committed user bases that might be leveraged for data[236]—are better positioned to exploit collapse. Most American users of these products are left with only a thicket of boilerplate that does not speak to the broader social impact of these agreements.[237]

To be sure, there are open technical questions, particularly the question of how much access to ever-more data will continue to contribute to AI

---

[233] *See supra* note 119 and sources cited therein (providing working definition of "agentic" AI).

[234] *See* Michelle Castillo, *Perplexity*, CNBC DISRUPTOR 50 (June 10, 2025 6:45 AM ET), https://www.cnbc.com/2025/06/10/perplexity-cnbc-disruptor-50.html ("Built by alumni from OpenAI, Meta, and Quora, Perplexity AI is attempting to create the next generation of search engines by combining generative AI with the internet.").

[235] Maxwell Zeff, *Perplexity Launches Comet, an AI-Powered Web Browser*, TECHCRUNCH (July 9, 2025, 8:00 AM), https://techcrunch.com/2025/07/09/perplexity-launches-comet-an-ai-powered-web-browser/ (discussing launch of "Comet" search product and quoting Perplexity CEO Aravind Srinivas: "Srinivas said in March that his goal with Comet was to 'develop an operating system with which you can do almost everything,' enabling Perplexity's AI to help users across apps and websites. Becoming the default browser for users can translate to 'infinite retention,' Srinivas said in June, which would ostensibly lead to more requests on Perplexity").

[236] For example, there have been reports that Amazon, which has troves of data from its customers, is developing its own chatbot, "Metis," to compete with the likes of Chat-GPT and Gemini. *See* Britney Nguyen, *Amazon is Working on A ChatGPT Competitor*, QUARTZ (Aug. 6, 2024), https://qz.com/amazon-generative-ai-chatbot-chatgpt-metis-1851558879. Amazon previously released Rufus, a "generative AI-powered conversational shopping assistant," for all U.S. customers. Rajiv Mehta, *How Customers are Making More Informed Shopping Decisions with Rufus, Amazon's Generative AI-Powered Shopping Assistant*, AMAZON (Sep. 18, 2024), https://www.aboutamazon.com/news/retail/how-to-use-amazon-rufus.

[237] *See* Andrew Keane Woods, *The New Social Contracts*, 77 VAND. L. REV. 1831, 1833 (2024) ("If ever there were a case for scrutinizing the public impact of private dealmaking, today's platform contracts are it; these are contracts of unique societal impact.").

development.[238] But for now, a large user base seems like a necessary, if not sufficient, condition for an "ask," with privacy policies and terms of service acting as the instruments to acquire data for AI systems. Furthermore, an "ask" that allows a company to acquire specific user datasets that are not available to other actors may be an especially valuable form of domain exploitation. That is because unique, proprietary datasets may position a company to fine-tune and adapt base models trained on vast corpuses of scraped data.[239] Data that a company acquires from its own users might allow that company to gain a competitive edge over other companies.[240] Any such competitive edge could further entrench that company as a leading AI developer—compounding the future impact on its users. [241]

---

[238] This point is a version of what technologists refer to as "scaling laws," or the idea that, "as we increase model size, training compute, and dataset size, language models get 'better.'" Narayanan & Kapoor, *AI Scaling Myths*, *supra* note 103.

[239] For instance, proprietary datasets can enable supervised fine-tuning that improves the performance of a deployed AI system. *See* Supervised Fine-Tuning, OpenAI Platform, https://platform.openai.com/docs/guides/supervised-fine-tuning (last visited Dec. 8, 2025) ("Supervised fine-tuning (SFT) lets you train an OpenAI model with examples for your specific use case."). In general, as of this writing, leading guides to fine-tuning focus on the need for task-specific examples and data. *See LLMs: Fine-Tuning, Distillation, and Prompt Engineering*, Google ML Crash Course (last updated Dec. 1, 2025), https://developers.google.com/machine-learning/crash-course/llm/tuning ("Fine-tuning trains on examples specific to the task your application will perform."); Model Fine-Tuning Concepts, Windows App Development (last updated Nov. 11, 2025), https://learn.microsoft.com/en-us/windows/ai/fine-tuning ("Fine-tuning helps you adapt pre-trained AI models to work better with your specific data and use cases. This technique can improve model performance while requiring less training data than building a model from scratch."); Yanyan Zhang et al., *Best Practices and Lessons For Fine-Tuning Anthropic's Claude 3 Haiku on Amazon Bedroc*k, AWS BLOGS (Nov. 1, 2024), https://aws.amazon.com/blogs/machine-learning/best-practices-and-lessons-for-fine-tuning-anthropics-claude-3-haiku-on-amazon-bedrock/ ("By fine-tuning, the LLM can adapt its knowledge base to specific data and tasks, resulting in enhanced task-specific capabilities."). *See also* Paul Ohm, *Focusing on Fine-Tuning: Understanding the Four Pathways for Shaping Generative AI*, 25 COLUM. SCI. & TECH. L. REV. 214 (2024) (discussing fine-tuning and its importance for generative AI development).

[240] *See, e.g.*, Thibault Schrepel & Alex 'Sandy' Pentland, *Competition Between AI Foundation Models: Dynamics and Policy Recommendations*, 34 INDUST. & CORP. CHANGE 1085, 1089 (2025) ("[A]ccess to unique data sets is critical."); Shayne Longpre et al., *A Large-Scale Audit of Dataset Licensing and Attribution in AI*, 6 NATURE MACHINE INTELL. 975, 976 (2024) ("We find a sharp and widening divide between commercially open and closed data, with the latter monopolizing more diverse and creative sources."). *Cf.* Daniel L. Rubinfeld & Michal S. Gal, *Access Barriers to Big Data*, 59 ARIZ. L. REV. 335, 350–51 (2017) (analyzing barriers to the collection of "big data" and arguing, even before the rise of generative AI models, that "unique access points to unique data may lead to situations in which the data cannot be easily replicated").

[241] I use the verb "could," not "will," because the outcome depends on both future technological developments and regulatory constraints. *See infra* text accompanying notes 244–249.

* * *

The status quo, in sum, risks entrenching contemporary balances of technological, social, and economic power. It does not prioritize the interests of most individuals, and it is likely to allow companies such as Meta, Alphabet (the parent company of Google), and Amazon to reproduce their historic commercial dominance in AI. In addition to obvious antitrust and competition law concerns about market concentration and the power of big tech companies in society, this result threatens public legitimacy and accountability in a constitutional democracy[242] and should worry even those who have entrenched power today.[243]

Admittedly, some caveats are in order. For one, the status quo is not entirely fixed. The conventional circle of power might expand slightly to include a select number of new AI leaders with popular AI products, such as OpenAI (which has received considerable funding from Microsoft[244]) and Anthropic (which has received computing support from both Google and Amazon[245]), and there might be new opportunities for players whose prominence faded in the early internet era, such as Microsoft[246] (in no small part thanks to its early partnership with OpenAI[247]). For another, open-source models like the one offered by the Chinese

---

[242] See discussion *infra* Part III.B.

[243] For an argument "against entrenchment" in a constitutional democracy, see Martha Minow, *Cooperation and Resistance to Entrenched Power: Some Preconditions for a Constitutional Democracy*, 63 DUQ. L. REV. 315, 329–31 (2025) ("To enable self-government and to protect minority groups and views, a constitutional democracy must promise to prevent a limited group of individuals from persistently controlling the government and other sources of power.").

[244] *See Microsoft and OpenAI Extend Partnership*, OFF. MICROSOFT BLOG (Jan. 23, 2023), https://blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership/; Cade Metz & Karen Weise, *Microsoft to Invest $10 Billion in OpenAI, the Creator of ChatGPT*, N.Y. TIMES (Jan. 23, 2023), https://www.nytimes.com/2023/01/23/business/microsoft-chatgpt-artificial-intelligence.html.

[245] *See* Gerrit De Vynck, *How Big Tech Is Co-Opting the Rising Stars of Artificial Intelligence*, WASH. POST. (Oct. 2, 2023), https://www.washingtonpost.com/technology/2023/09/30/anthropic-amazon-artificial-intelligence/.

[246] Microsoft's economic fortunes have soared with the rise of ChatGPT and generative AI systems. *See, e.g.*, Karen Weise & Cade Metz, *How Microsoft's Satya Nadella Became Tech's Steely Eyed A.I. Gambler*, N.Y. TIMES (July 16, 2024), https://www.nytimes.com/2024/07/14/technology/microsoft-ai-satya-nadella.html (reporting that Microsoft's sizeable AI investments have "pushed Microsoft's worth up 70 percent to more than $3.3 trillion, making Microsoft one of three companies (with the chip maker Nvidia, another A.I. star, and Apple) vying to be the most valuable publicly traded company in the world").

[247] *See* BRIAN MERCHANT, A.I. NOW INST., AI GENERATED BUSINESS 25-26 (2024), https://ainowinstitute.org/wp-content/uploads/2024/12/AI-Now_Generative-AI-Business-Models.pdf (describing benefits to Microsoft from its partnership with OpenAI).

company DeepSeek might further complicate the picture.[248] But so long as access to massive amounts of data remains essential for AI developers to create cutting-edge models in the first place, entities that have an edge in data acquisition will have an edge in AI development.[249] Users, as well as non-famous, non-industry content creators, will bear a disproportionate cost. The distributive inequities of the status quo must factor into the overall evaluation of collapse and its consequences. Furthermore, as the next Part contends, there are also steep costs for the integrity of law itself.

B. The Governance Toll of Collapse: Law's Legibility and Legitimacy

This Part argues that doctrinal collapse threatens the rule of law. At least some readers might chafe at this claim, on the grounds that rule of law concerns arise only for public actors, whereas collapse in AI development involves private companies.[250] I adopt a more capacious understanding of the rule of law, with an eye to the ways that inter-regime doctrinal collapse weakens the legal system's

[248] *See* Will Douglas Haven, *How DeepSeek Ripped Up the AI Playbook—And Why Everyone's Going to Follow Its Lead*, MIT TECH. REV. (Jan. 31, 2025), https://www.technologyreview.com/2025/01/31/1110740/how-deepseek-ripped-up-the-ai-playbook-and-why-everyones-going-to-follow-it/.

[249] Notably, DeepSeek's innovations do not appear to have changed this requirement. Its model has distinct compute requirements, but the data needs appear unchanged—and indeed, part of DeepSeek's competitive advantage comes from lower data acquisition costs. *Id.* ("[DeepSeek] has also found cheaper ways to create large data sets."). In addition, DeepSeek itself is arguably one of a select group of players capable of producing the underlying open-source model. *Cf.* Jeffrey Ding, ChinAI #298: A Rejoinder on DeepSeek and export controls, CHINAI NEWSLETTER (Feb. 3, 2025), https://chinai.substack.com/ (providing translation of Chinese-language document: ". . . [T]here are only a few companies in the world that have enough resources and data to touch the ceiling of Scaling law. Because first of all, it requires sufficiently powerful infrastructure and computing resources, and secondly, it requires sufficient training data.").

[250] Thank you to Aziz Huq for helpful questions and comments on this point. On the conventional scope of debates about the rule of law, as well as analysis of "the possibility that peripheral cases involve capricious private power," see AZIZ Z. HUQ, THE RULE OF LAW: A VERY SHORT INTRODUCTION 10–11 (2024).

capacity to amply constrain the arbitrary exercise of power in our constitutional democracy,[251] whether that power is in the hands of state or non-state actors.[252]

Recognizing that the rule of law is a multi-faceted concept with formal, procedural, and substantive varieties, this Article embraces a thick, formal understanding.[253] By thick, I mean that merely setting out rules, such that there is "thin" rule *by* law, will not produce the conditions that the rule of law demands.[254] By "formal," I reference Lon Fuller's contention that a system of law-making requires more than a clear set of rules.[255] Among other criteria that Professor Fuller sets out in his canonical account, in a rule-of-law system, the rules must be publicly understandable and applied in a consistent, justifiable manner.[256] There is, to be sure, disagreement in the literature about Professor Fuller's formalist understanding of what is necessary for a system of law-making to comport with the rule of law, and what that means for the system of law itself.[257] These disagreements are the subject of a vast jurisprudential literature,

---

[251] *See* Martin Krygier, *What's the Point of the Rule of Law?*, 67 BUFF. L. REV. 743, 760–61 (2019); Martin Krygier, *The Rule of Law: Pasts, Presents, and Two Possible Futures*, 12 ANN. REV. L. & SOC. SCI. 199, 204 (2016). *Cf.* COHEN, BETWEEN TRUTH AND POWER, *supra* note 25, at 204 ("If the new network-and-standard-based governance institutions are to serve the overarching institutional functions that traditionally have informed thicker versions of rule-of-law thinking—functions that, to borrow Martin Krygier's formulation, temper the arbitrary exercise of power—both institutions and constructs will need to adapt" (citing Krygier, *The Rule of Law: Pasts, Presents, and Two Possible Futures*, *supra* note 251)).

[252] *See* Martin Krygier, *The Rule of Law: Legality, Teleology, Sociology*, *in* RELOCATING THE RULE OF LAW 45, 59 (Gianluigi Palombella & Neil Walker eds., 2009); Krygier, *The Rule of Law: Pasts, Presents, and Two Possible Futures*, *supra* note 251, at 203; *see also* Solow-Niederman, *Algorithmic Grey Holes*, 5 J. L. & INNOV. 116, 134 n.63 (2023) ("This Essay focuses exclusively on public actors, where the rule of law connection is most explicit. But the phenomena identified here may sweep beyond state action."). Even if one does not agree that this rule of law analysis ought to extend to private actors, the rise of a tech "oligarchy" calls for attention to the relationship between market power and political power. In particular, to the extent that one is concerned that there is a tech oligarchy, it becomes even more important to recognize how private actors can take advantage of doctrinal collapse in ways that affect the broader legal system. For "an account of oligarchy and, more specifically, of tech oligarchy within contemporary political economy," see Julie E. Cohen, *Oligarchy, State, and Cryptopia*, 94 FORDHAM L. REV. 563, 567 (2025).

[253] For an overview of formal, substantive, and procedural versions of rule of law, see Aziz Z. Huq, *Artificial Intelligence and the Rule of Law*, *in* THE ROUTLEDGE HANDBOOK OF THE RULE OF LAW 260, 265 (Sevel, ed. 2024).

[254] *See* Alicia G. Solow-Niederman, *Algorithmic Grey Holes*, 5 J. L. & INNOV. 116, 120 (2022) ("A thicker understanding might call for rule *of* law, in the sense of requiring the system of law-making to comport with a broader set of legal principles.")

[255] *See* LON L. FULLER, THE MORALITY OF LAW 38-39 (rev. 1964).

[256] *See id*. Professor Fuller set forth eight criteria: laws must be general, open, prospective, clear, consistent, capable of being obeyed, stable, and upheld by officials. *See id.*

[257] *See, e.g.*, Jeremy Waldron, *Getting to the Rule of Law: The Rule of Law and the Importance of Procedure*, 50 NOMOS 3, 5-6 (James E. Fleming ed. 2011) (arguing that "laundry lists of demands," such as Professor Fuller's eight formal principles, must be accompanied by "a list of procedural characteristics that are equally indispensable"); Jeremy Waldron, *Positivism and*

and rightly so. However, even if one does not agree with Professor Fuller's articulation, it's hard to see how a system of law, understood to involve actions by both state and non-state actors, can be considered non-arbitrary if the public cannot know or predict how it will apply.

Doctrinal collapse threatens the rule of law to the extent that it makes the application of law unpredictable, internally inconsistent, or otherwise arbitrary. More precisely, inter-regime doctrinal collapse undermines the rule of law when it enables exploitation by sophisticated actors, which undercuts the legal system's ability to govern legal claims and regulate conduct in non-arbitrary ways. This result is not limited to AI and data acquisition. Because inter-regime doctrinal collapse always involves blurring of distinct and fundamentally irreconcilable animating logics, it is at odds with crisp, consistent, and predictable legal lines. When sophisticated actors exploit collapse, they are manipulating the instability of the two domains.

This manipulation of doctrinal instability can have broader costs for the legal system, particularly when it comes to public accountability and legitimacy.[258] Recall that companies focus on the "public" nature of data to make privacy law arguments; copyright law arguments; or both.[259] Moreover, the current internal structures of copyright and privacy law permit a company to marshal conflicting claims at different points in time.[260] At the outset of litigation, a firm like OpenAI may focus on the "public" nature of data both to defend itself from copyright

Legality: Hart's Equivocal Response to Fuller, 83 N.Y.U. L. REV. 1135, 1135-38 (2008) (discussing disagreement between H.L.A. Hart and Professor Fuller); Ronald Dworkin, *Philosophy, Morality and Law: Observations Prompted by Professor Fuller's Novel Claim*, 113 U. PA. L. REV. 668, 668 (1965) ("I take Fuller's recent book, 'The Morality of Law,' to be an unsuccessful attempt to establish a novel claim about law and morality.").

[258] My emphasis is the application of traditional forms of law (here, copyright law and privacy law) and associated social norms to data acquisition and AI systems. This Article is thus distinct from a body of work on the use of a technology (such as blockchain) to implement legal mandates, and the rule of law implications of such "code-driven" or "data-driven" law. For selected sources in this literature, see, for example, LAURENCE E. DIVER, DIGISPRUDENCE: CODE AS LAW REBOOTED (2021); Mireille Hildebrandt, *Algorithmic Regulation and the Rule of Law*, 376 Phil. Trans. R. Soc. A 1 (2018); Roger Brownsword, *Technological Management and the Rule of Law*, 8 L., INNOV. & TECH. 100 (2016); Marco Goldoni, *The Politics of Code as Law: Toward Input Reasons*, *in Information and Law in* TRANSITION: FREEDOM OF SPEECH, THE INTERNET, PRIVACY AND DEMOCRACY IN THE 21ST CENTURY (Anna-Sara Lind, Jane Reichel, & Inger Österdahl eds., 2015); Lodewijk Asscher, *'Code' As Law - Using Fuller to Assess Code Rules, in* CODING REGULATION: ESSAYS ON THE NORMATIVE ROLE OF INFORMATION TECHNOLOGY 61 (Egbert Dommering & Lodewijk Asscher eds., 2006). Much of this work builds on the foundational scholarship of Lawrence Lessig and Joel Reidenberg. *See generally* LESSIG, *supra* note 125 (arguing that "code is law"); Reidenberg, *supra* note 125 (introducing the idea of "Lex Informatica" to refer to "the set of rules for information flows imposed by technology and communication networks").

[259] See discussion *supra* Part II.B.

[260] See discussion *supra* Part II.B.

costs and to contend that there is no privacy interest in the data used to train its AI model. But the same firm may simultaneously refuse to disclose the training data and claim that the dataset is private or proprietary. In theory, each regime features distinct doctrinal and normative lines. In practice, it is extraordinarily confusing to trace how the law does or should apply when a company claims that the same regulatory object (data) is both public (and open to the company to acquire) and closed off from public access. The boundaries of the regimes blur, which creates confusion about what each regime requires and also creates opportunities for further arguments that destabilize both regimes even more.[261] Notably, this confusion and unsettlement can begin with private parties' public-facing rhetoric (creating conceptual blurring) and / or litigation stances (creating functional blurring) regardless of whether these arguments are ever accepted by a court in a final opinion (which might create institutional blurring).[262]

Such collapse has rule of law costs because the regulatory function of each domain becomes unclear and arbitrary as privacy is strategically minimized, then maximized, in ways that do not cohere over time. On the one hand, as discussed in Part III, privacy-based claims generally do not feature prominently in AI firms' public statements or legal briefs about their training data,[263] and both the "buy" and "ask" tactics discount user privacy interests once data is disclosed to a platform.[264] Moreover, when it serves them to do so, AI developers further minimize privacy interests. For instance, a developer might highlight individuals' voluntary disclosure of information to the company, underscoring the users' choice to accept the platform's privacy policy and terms of service. In one pending lawsuit, for example, Google rejects privacy objections, arguing that "[t]here is no basis for Plaintiffs to refuse to identify other email addresses they already supplied to Google when signing up for its services."[265] The company continues: "Google . . . will investigate what services Plaintiffs used and whether Plaintiffs uploaded their copyrighted works to the services. This too raises no 'privacy' concerns; any works Plaintiffs upload to Google's services have also already been voluntarily provided (and licensed) to Google."[266] In this and similar instances, privacy interests are minimized.

On the other hand, when user privacy can be strategically invoked to resist discovery, it suddenly becomes a leading argument. Indeed, claims about the need to protect "user privacy" for information shared with the platform during

---

[261] See discussion *supra* Parts I–II.

[262] See discussion *supra* text accompanying notes 46-47.

[263] See discussion *supra* Part II.B.

[264] See discussion *supra* Part II.D.

[265] Discovery Letter Brief at 4, In re Google Generative A.I. Copyright Litig., No. 5:23-cv-03440 (N.D. Cal. May 8, 2025), ECF No. 134.

[266] Discovery Letter Brief at 4, In re Google Generative A.I. Copyright Litig., No. 5:23-cv-03440 (N.D. Cal. May 8, 2025), ECF No. 134.

use feature in multiple generative AI lawsuits. For instance, in *Concord v. Anthropic*, Anthropic argues that the records sought "contain private and sensitive information" and that "[p]ublicizing these records risks violating Anthropic users' privacy and undermines Anthropic's commitment to keep its users' information confidential."[267] And in *New York Times v. OpenAI*, OpenAI argues that the court's order to preserve all output logs "requir[es] OpenAI to disregard the privacy interests of its users [in a way that] is wholly disproportionate to the needs of the case and unwarranted"[268] and repeatedly invokes user privacy to contest disclosure of user chat logs to the plaintiffs in both litigation[269] and in the court of public opinion.[270] In these and similar instances, privacy interests are suddenly maximized—even though they were minimized throughout training, deployment, and earlier court filings and public statements.

---

[267] Joint Discovery Dispute Statement Regarding Publishers' Challenges to Anthropic's Confidentiality Designations at 7-8, Concord Music Grp., Inc. v. Anthropic PBC, No. 5:24-cv-03811 (N.D. Cal. May 29, 2025), ECF No. 380; Joint Discovery Dispute Statement Regarding Publishers' Challenges to Anthropic's Confidentiality Designations at 8-9, Concord Music Grp., Inc. v. Anthropic PBC, No. 5:24-cv-03811 (N.D. Cal. June 26, 2024), ECF No. 345.

[268] Reconsideration of Order Directing OpenAI to Preserve Output Logs; In Re: OpenAI, Inc., Copyright Infringement Litigation, No. 1:25-md-3143, at 2, (S.D.N.Y. May 15, 2025), ECF No. 40 (document related to NYT v. Microsoft Corporation, et al., No. 1:23-cv-11195).

[269] *See* Letter to Magistrate Judge Ona T. Wang, Re: OpenAI, Inc., Copyright Infringement Litigation, No. 1:25-md-3143, at 1–3, (S.D.N.Y. Oct. 30, 2025), ECF No. 717 (document related to NYT v. Microsoft Corporation, et al., No. 1:23-cv-11195); Letter to Magistrate Judge Ona T. Wang, Re: OpenAI, Inc., Copyright Infringement Litigation, No. 1:25-md-3143, at 1–3, (S.D.N.Y. Oct. 30, 2025), ECF No. 742 (document related to NYT v. Microsoft Corporation, et al., No. 1:23-cv-11195) (seeking reconsideration after Magistrate Judge Wang's order granting News Plaintiffs' motion to compel and directing OpenAI to produce the 20 million de-identified user chat logs at issue, *see id.* at ECF No. 734, and asserting that the chat log data at issue "belongs to ChatGPT users all over the world—families, students, teachers, government officials, financial analysts, programmers, lawyers, doctors, therapists, and even journalists—whose private thoughts and confidential business information may now be exposed in this lawsuit"); Letter to Magistrate Judge Ona T. Wang, Re: OpenAI, Inc., Copyright Infringement Litigation, No. 1:25-md-3143, at 1–2, (S.D.N.Y. Nov. 14, 2025), ECF No. 752 (document related to NYT v. Microsoft Corporation, et al., No. 1:23-cv-11195) (explaining that the agreed-upon deidentification process will not resolve all privacy concerns and again emphasizing "the privacy interests of millions of ChatGPT users worldwide"); Updated Memorandum of Law in Support of OpenAI's Rule 72(A) Objections to the Orders Compelling Production of ChatGPT Conversation Logs at MDL ECF 734, MDF ECF 896, and MDL 910, Re: OpenAI, Inc., Copyright Infringement Litigation, No. 1:25-md-3143, at 1–2, 5–12 (S.D.N.Y. Dec. 12, 2025), ECF No. 935 (objecting to Magistrate Judge Wang's denial of OpenAI's motion for reconsideration, *see id.* at ECF No. 896, as clearly erroneous and contending that the Magistrate Judge's "Order failed to adequately consider the privacy interests of absent non-parties").

[270] *See Fighting the New York Times' Invasion of User Privacy*, OPENAI (Nov. 12, 2025), https://openai.com/index/fighting-nyt-user-privacy-invasion/.

Confused about which doctrine controls, and why? That's the point. No matter the validity of any underlying copyright and/or privacy argument, this sort of strategic toggling between regimes (and shifting stances with respect to when an issue, such as privacy, is even relevant) has consequences for the rule of law.[271] Some might say, that's just good lawyering. Perhaps. But at what cost for the system of law? When private actors can choose which doctrines apply, at which times, depending on what serves them, controlling legal regimes lose their consistency and coherence. This outcome can make legal rules less publicly accessible and less easily understandable by members of the public.[272] Put simply, private actors' toggling can make the law increasingly illegible. Even if courts check this risk and manage collapse—a possibility that Part V considers—there are still rule of law costs because the public cannot predict which rules will apply, at which point. Copyright law, for instance, is invoked both to deny property claims in training data (because it is "public"), and later to shield the company's data, closely followed by other IP claims that the resulting models as proprietary and protected. Privacy, too, is minimized when it threatens access to training data, yet later maximized as grounds for non-disclosure of user data. That alone makes the force of law seem arbitrary and compromises the public accountability and legitimacy of law itself.

To be sure, not all of this is new. Scholars have long contested the "pretextual" invocation of privacy and lamented privacy law's "weaponization" and "cooptation."[273] For example, companies have previously made arguments that Rory Van Loo labels as "privacy pretexts."[274] Professor Van Loo has exposed how, in the competition law and accountability context, companies "cit[e] privacy to advance their interests at the expense of individuals."[275] What has not yet received adequate attention, though, is how the structure of overlapping legal regimes facilitates pretextual arguments like these. This

[271] One can think, as I do, that there *are* in fact serious privacy concerns here, yet still object to such strategic toggling. Indeed, that is the point: It's vital to focus not only on any singular field of law, but also on the rule of law costs across domains.

[272] Again, I contend that there are rule of law costs even if courts do not accept the arguments, which would signal institutional blurring. *See supra* text accompanying note 261.

[273] On the weaponization of privacy, see, for example, COHEN, BETWEEN TRUTH AND POWER, *supra* note 25, at 46-47 (arguing that "informational capitalism" has permitted informational property rights to be restructured and reworked through the "self-interested, strategic activities of many different players"); ARI EZRA WALDMAN, INDUSTRY UNBOUND 7 (2021) (arguing that companies have watered down privacy protections into largely symbolic compliance procedures). On the "pretextual" use of privacy arguments in other contexts, see Rory Van Loo, *Privacy Pretexts*, 108 CORNELL L. REV. 1, 2-4 (2022); Susan Hazeldean, *Privacy as Pretext*, 104 CORNELL L. REV. 1719, 1721 (2019); Christina Koningisor, *Coopting Privacy*, 105 B.U. L. REV. 765, 771-73 (2025). For a summary of recent work, followed by an analysis of pretextual privacy arguments in the GDPR context, see Neil Richards, *The GDPR as Privacy Pretext and the Problem of Co-Opting Privacy*, 73 HASTINGS L.J. 1511, 1514 (2022).

[274] Van Loo, *supra* note 273, at 5-6.

[275] Van Loo, *supra* note 273, at 3.

Article's account of inter-regime doctrinal collapse directs attention to this enabling dynamic.

Furthermore, the inconsistent and unpredictable application of rules across the copyright law-privacy law boundary threatens the rule of law because it obscures the underlying justification for applying a particular doctrine. This outcome is more than legal ambiguity within a domain or manageable doctrinal tension across domains. Law is of course never entirely certain, and it can never be complete.[276] Legal systems can manage uncertainty, which operates within established frameworks. Collapse destabilizes the frame. Inter-regime doctrinal collapse makes it much harder to trace the animating logic of a particular domain, as applied to a particular set of facts. Admittedly, understood this way, any legal argument that involves overlapping domains with different animating logics may come with a rule of law cost, at least to some extent. The question is whether that cost is worth paying, given what this legal instability allows companies to construct. For data acquisition and AI, if one believes that the contemporary political economy of AI development is desirable, then perhaps the current path of AI produces a net social benefit. But if one thinks that there are inequitable results, as I do, then the rule of law cost creates additional reasons to focus on doctrinal collapse and the exploitation that it enables.

* * *

Regardless of its potential to spur technological or regulatory innovation, the contemporary political economic and jurisprudential costs of collapse in data acquisition are just too high. If the collapse of copyright law and privacy law goes unchecked, law risks losing the capacity to operate in a consistent way, and the public risks losing faith that law can constrain power. The next Part considers what to do about it.

IV. Reckoning with Collapse

In the face of collapse, the goal is not to restore perfect clarity or to impose artificial doctrinal lines. Inter-regime doctrinal collapse does not itself have a valence, and some degree of overlap and blurring between regimes is inevitable.[277] It can even be good. Collapse becomes a problem, though, when it disproportionately empowers privileged actors and corrodes law's capacity to govern.[278] The best response is to understand when, where, and why collapse is happening. This analysis enables judges, regulators, and scholars to take steps

[276] Pistor, *supra* note 51, at 251 ("The concept of incompleteness recognizes that the future is unknown and inherently unknowable, a notion that Frank Knight deemed fundamental uncertainty." (citing FRANK KNIGHT, RISK, UNCERTAINTY, AND PROFIT 43-44, 46 (1921))).

[277] See discussion *supra* text accompanying notes 3–9 and Part I.

[278] See discussion *supra* Part III.

that recalibrate law's capacity to meaningfully constrain private power in a transformed social and technological environment.

The remainder of this Article traces a path to that recalibration, with an eye to marking out productive options rather than dictating any one way forward. My goal is to permit conversation about collapse and its consequences, and to highlight the points that I find most problematic—but I intend this conversation to be a generative one and not a preemptive one. To that end, Part V.A emphasizes the value of identifying collapse and suggests how this recognition can be a useful step for policymakers and advocates. Part V.B then considers two sets of potential reforms: The first path is a more incremental approach that works within the structure of law as it is to mitigate the worst outcomes of collapse, and the second path is a more interventionist approach that would rework the structure of law itself, with an eye to preventing problematic results in the first instance.

A. Recognition

For legal institutions to manage inter-domain conflict in principled ways, the first step is to recognize collapse. Collapse can be diagnostic. The indicators of collapse identified in this Article (overlap-blurring-irreconcilability[279]) signal that a particular regulatory object does not naturally fit within a settled doctrine or normative consensus. Recognizing collapse thus helps scholars and advocates pinpoint which regulatory objects are likely to be focal points of contestation. This recognition can promote constructive "tussles" that expose underlying arguments about values and how to regulate.[280] When these tussles come to the surface, it is less likely that one objective will be prioritized without interrogating whether this is an optimal outcome or recognizing tradeoffs with other legal and social values.[281]

This recognition is not only theoretically useful, but also quite practical. It can strengthen arguments by those who seek to challenge the weaponization or manipulation of a complex and contested concept. More generally, it facilitates contestation of the power dynamics enabled by the legal status quo. For instance, in addition to discussing copyright and IP interests solely in the focused context of specific lawsuits, this Article's theory of inter-domain doctrinal collapse suggests that civil society groups might seek to create coalitions that bridge copyright and privacy law interests. As one concrete example, chatbot users who did not realize that their conversation would be shared publicly when they hit

[279] See discussion *supra* Part I.A.

[280] *See* Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-By-Design*, 106 CAL. L. REV. 697, 743-45 (2018); David D. Clark et al., *Tussle in Cyberspace: Defining Tomorrow's Internet*, 13 IEEE/ACM TRANSACTIONS ON NETWORKING 462, 466 (2005).

[281] *See* Mulligan & Bamberger, *supra* note 280, at 743-45; Clark et al., *supra* note 280, at 466.

"share" and artists whose work has been used to train AI models are objecting to different things, yet both groups are arguably exploited by the AI development status quo (for better or worse). Especially because the form of exploitation that collapse enables involves manipulation of both legal and social costs,[282] social movements that cross both copyright and privacy might shift the political economy of data acquisition. Whether pursuing any such interest convergence is, on balance, beneficial to all parties is a question that warrants further research.[283] This Article's theorization of inter-regime doctrinal collapse facilitates that investigation and, when appropriate, can help civil society groups to use overlap and blurring as diagnostic signals that there may be space to build bridges.

Although valuable, responses of this sort do not suffice for principled management of collapse and its impact on both individuals and on the rule of law. The next question is how to design governance institutions that account for the negative individual and system-wide outcomes that collapse can enable. Again, the goal is not to "solve" collapse. Because collapse is an enabling condition and not an end point, that would be a category error. The goal is to reduce the possibility that collapse leads to harmful exploitation or corrodes law's legibility and legitimacy in ways that disserve the public interest. The most auspicious possibilities draw on existing legal resources. The remainder of this Part sketches two potential paths forward—one more small-c conservative and incremental, and one more reform-minded—with the goal of seeding future conversations.

B. Response

1. The Incremental Path: Adapting Existing Legal Structures

The more conservative path is to accept that, at least for AI development, inter-regime doctrinal collapse and subsequent exploitation by powerful private actors is inevitable, and yet still take steps to minimize harmful outcomes. Notably, law has developed entire fields to manage and reconcile competing claims across jurisdictions, whether vertically (as between federal and state

[282] See discussion *supra* Part I.

[283] *See generally* Derrick A. Bell, Jr., Brown v. Board of Education *and the Interest-Convergence Dilemma*, 93 HARV. L. REV. 518 (1980) (providing canonical account of interest convergence). *See also* Danielle Keats Citron, 89 B.U. L. REV 61, 84-86 (2009) (invoking and applying Bell's theory in the context of cyber civil rights).

sovereigns)[284] or horizontally (as between co-equal state sovereigns).[285] But the law lacks an analogous set of resources to manage inter-regime conflict between competing doctrines, rather than competing sovereigns. To be clear, this is an analogy. I do not argue that doctrinal collapse should be treated as a literal conflict-of-laws problem. Still, the analytic challenge is similar: How can we create principles to govern when multiple legal domains apply to the same terrain, yet point in conflicting and/or normatively inconsistent ways?

A conflict-of-laws inspired response of the sort proposed in this Article should be tailored to specific institutional actors. Because conflict of laws involves legal arguments (as opposed to public rhetoric), the courts are best suited to implement this form of collapse management. For example, if a court is resolving a conflict that involves competing copyright law and privacy law interests and arguments, and functional blurring of the two domains, then the judge might insist on a rebuttable "anti-switching" presumption. As a rough cut, the idea is that a party in litigation cannot assert mutually incompatible claims at different points in the lawsuit, absent a sufficiently compelling reason to defeat the presumption.

This presumption would speak directly to many of the copyright and privacy conflicts discussed above. For instance, it would make it far harder for companies developing AI to strategically move between copyright and privacy law claims about "public" data. Such an intervention would promote internal doctrinal consistency and rule of law coherence. Additionally, this presumption would simultaneously constrain a sophisticated actor's ability to exploit the law in self-serving ways and thereby curb political economic disparities. Significantly, because it is a rebuttable presumption, this intervention would require the court to engage in case-by-case, fact specific analysis; thus, it would avoid locking in any assumption that copyright law or privacy law should always take priority and promote a focused analysis of the interests as they are presented in a particular legal dispute.[286]

Courts faced with conflicting legal regimes might also more directly weigh the competing normative values implicated by each body of law. For instance, returning to the case study in copyright law and privacy law once more, a court might assess whether copyright law's incentive-based, property logic, or privacy

[284] *See* Paul S. Berman , Roey Goldstein & Sophie Leff, *Conflicts of Law and the Abortion War Between the States*, 172 U. PA. L. REV. 399, 455–56 (2024). *See also* Murphy v. Nat'l Collegiate Athletic Ass'n, 584 U.S. 453, 470–72, 477–79 (2018) (discussing Supremacy Clause, anti-commandeering principle, and preemption doctrine).

[285] *See* Restatement (Second) of Conflict of Laws §§ 2 & cmt. a, § 6 & cmt. on subsection (2) (A.L.I. 2025).

[286] On the potential benefits of AI litigation as regulation, including fact-intensive, precise analysis of this sort, see Alicia Solow-Niederman, *Do Cases Generate Bad AI Law?*, 25 COLUM. SCI. & TECH. L. REV. 261, 275-78 (2024), and sources cited therein.

law's control logic is the better fit. As an example of what this would look like in practice, consider one recent district court case involving data scraping in a non-AI context, *X Corp. v. Bright Data*. Although the court rejected the defendant's state-law claims as preempted by the Copyright Act, the *Bright Data* Court made clear that its holding would not automatically extend to cases that more squarely presented privacy questions: "It does not follow, however, that state law [ ] interests are inevitably preempted whenever their recognition would burden the enjoyment of the benefits of copyright."[287] The *Bright Data* Court continued with an illustrative example that specified how particular legal claims related to distinct normative interests in each legal regime: "[T]he Copyright Act should not preempt analogous state-law claims asserted by a social media company to protect its users' privacy because the protection of privacy is not a function of the copyright law, which offers a limited monopoly to encourage ultimate public access to the creative work of the author."[288] Thus, the court explicitly considered how data disputes implicate both copyright law and privacy law interests, taking care to cabin its ruling to maintain boundaries between the normative lines of the two regimes.

Critically, any such judicial analysis of the conflicting regimes would need to be highly fact-specific and would aim to assess which values have a stronger normative claim, on the facts and equities presented. There is admittedly a risk that such assessment amounts to improper judicial policymaking, rather than application of law to the facts.[289] Yet there are also potential benefits. For instance, connecting the facts and law of the case to these broader considerations about the normative basis of the doctrines might counteract some of the limitations inherent in private, individualized adjudication of AI controversies, which often has broader social impacts beyond the narrow case.[290] Future work can and should assess how to apportion any such judicial discretion in ways that are consistent with the role of the judge.

Careful analysis of the normative values implicated in a particular dispute might have other benefits, too. When litigants are aware that the court will conduct an analysis of the normative values associated with different doctrinal

[287] X Corp. v. Bright Data Ltd., 733 F. Supp. 3d 832, 850, 852-53 (N.D. Cal. 2024) (internal citations omitted); *see also* Henderson & Lemley, *supra* note 54, at 1365-67 (discussing Judge Alsup's analysis of copyright and privacy arguments in *Bright Data*). Because the present Article focuses on information privacy and copyright law, I reserve further analysis of inter-domain doctrinal collapse between copyright and contract law for future work. For discussion of recent cases involving data scraping, contract law claims, and copyright preemption arguments, see Henderson & Lemley, *supra* note 54, at 1357–65.

[288] Bright Data Ltd., 733 F. Supp. at 85.

[289] Thank you to Richard Re for raising this point.

[290] On the potential costs and risks of AI adjudication as regulation, emphasizing "concerns with the quality of a judicial decision itself and . . . concerns that the process of private adjudication will fall short of its deliberative potential," see Solow-Niederman, *Do Cases Generate Bad AI Law?*, *supra* note 286, at 278-82, and sources cited therein.

claims, there will be a natural incentive for parties to articulate the normative stakes of their arguments. There is admittedly a non-negligible risk that normative arguments of this sort will freeze in place the understanding of a legal regime that exists at a particular point in time, such as the contemporary understanding of copyright incentives and privacy as control.[291] But there's also the possibility that parties will produce fresh and creative arguments. What's more, if a party knows that they will need to explain the normative stakes of their argument, then there may be a natural bridge between scholarship that advances a particular normative perspective on what a contested legal domain like privacy or copyright law should be, and the arguments that parties make in court. Over time, this scholarly-adjudicative dialogue could inject new vitality into these contested domains. Discourse of this sort matters because judicial recognition and reasoning are not just reactive. Courts express important legal values and shape how domains cohere.[292]

2. The Reformist Path: Shifting the Structure of Law

Another path is more reform-minded, requiring regulatory interventions to change the structure of law by reducing inter-regime asymmetries and, in so doing, reducing the likelihood of problematic downstream exploitation.

The most direct path involves regulatory reforms to strengthen information privacy protections, which would change the overall legal structures that produce collapse and enable problematic forms of exploitation. Recall that structural weaknesses within one domain can create conditions for collapse of two partially overlapping domains.[293] This comparative structural vulnerability then facilitates exploitation across domains, contributing to especially problematic forms of further collapse. This Article's analysis reveals that weaknesses in the dominant implementation of a singular domain are not just a problem for that domain of law; rather, there are far-reaching consequences.

Privacy law is a case in point. For decades, scholars have critiqued the regulatory status quo. The critiques are multifaceted and could fill multiple law review volumes; to name but a few, concerns range from the deficiencies of "notice-and-choice,"[294] to the limitations of individual control,[295] to the lack of

---

[291] See discussion *supra* Part II.

[292] *See, e.g.*, Solow-Niederman, *Do Cases Generate Bad AI Law?*, *supra* note 286, at 276 ("[T]he very act of adjudicating can be instrumentally valuable for both individuals and for the public"); Danielle K. Citron, *Law's Expressive Value in Combating Cyber Gender Harassment*, 108 MICH. L. REV. 373, 407-413 (2009).

[293] *See* discussion *supra* text accompanying note 46.

[294] *See, e.g.*, Reidenberg et al., *supra* note 102, at 490–96 & nn. 17-44.

[295] *See, e.g.*, sources cited *supra* note 98.

attention to privacy's fundamentally relational nature,[296] to the regulatory regime's predominantly procedural, compliance oriented provisions and lack of substantive protections,[297] to the failure to amply account for the interests of marginalized groups.[298] The list goes on, but this Article is long enough, so let that suffice. With these critiques come a bevy of potential interventions.[299] My goal here is not to endorse any reform agenda, though I do think, given the limitations of an individualistic, control-centered model and the manipulation of the concept of "public" in AI development,[300] that interventions that speak to these points are auspicious starts.

The central takeaway, for purposes of this piece, is that even those who care less about privacy than the average privacy scholar should pay attention to longstanding weaknesses in the contemporary structure of information privacy protections, particularly at the U.S. federal level. Regulators and scholars alike must recognize that allowing weaknesses such as those in privacy law to persist comes with a structural cost that affects other areas of law and other social dynamics, too. I happen to think that privacy is well worth protecting on its own merits. But even if one does not agree, the United States' lack of overarching, substantive protections for information privacy at the federal level has implications that go far beyond privacy. When the boundaries of a weak domain (like privacy) can be too easily manipulated, the other partially overlapping domain (here, copyright) also loses its doctrinal and normative integrity.[301] And, at least when it comes to data, it is the technology "haves" who can take

---

[296] *See, e.g.*, sources cited *supra* note 98.

[297] *See, e.g.*, Ari Ezra Waldman, *Privacy, Practice, and Performance*, 110 CALIF. L. REV. 1221, 1225–26 (2022) ("[A]ll of [the privacy practices] are performative, and our acculturation to them has entrenched them and defined our relationship to, and assumptions about, privacy law."); Solon Barocas & Helen Nissenbaum, *Big Data's End Run Around Procedural Privacy Protections*, 57 COMM. ACM 31, 33 (2014) ("It is time to recognize the limits of purely procedural approaches to protecting privacy. It is time to confront the substantive values at stake in these information practices and to decide what choices can and cannot legitimately be placed before us—for our consent.").

[298] *See generally* DANIELLE KEATS CITRON, THE FIGHT FOR PRIVACY (2022) (exposing the disproportionate impact of intimate privacy invasions on women and marginalized groups and advocating a civil right to intimate privacy); *see also* Anita L. Allen, *Dismantling the "Black Opticon": Privacy, Race Equity, and Online Data-Protection Reform*, 131 YALE L.J. FORUM 907, (2022) (coining the term "Black Opticon" and critically assessing "African Americans' vulnerabilities to varied forms of discriminatory oversurveillance, exclusion, and fraud—aspects of which are shared by other historically enslaved and subordinated groups in the United States and worldwide"); SCOTT SKINNER-THOMPSON, PRIVACY AT THE MARGINS (2020) (emphasizing the importance of privacy protections for marginalized groups); Alvaro M. Bedoya, *Privacy as a Civil Right*, 50 N.M. L. REV. 301, 306 (2020) (arguing that privacy is not only a civil liberty, but also a civil right).

[299] Among many, many possible resources, see sources cited *supra* notes 294–297.

[300] See discussion *supra* Part II.

[301] See discussion *supra* Parts II-III.

advantage of the disarray.[302] Stronger privacy laws would make it harder for powerful actors to exploit conditions of collapse in their favor. And for that reason, they warrant serious attention.

In addition, a more subtle and challenging path is for scholars and policymakers to think differently about the relationship between partially overlapping, yet distinct, bodies of law that apply to the same regulatory object (as is the case for privacy law, copyright law, and data). Future work might investigate how to adapt the work of legal pluralists to understand sites of inter-regime doctrinal collapse as "hybrid legal spaces, where more than one legal, or quasi-legal, regime occupies the same social field."[303] The focus here is less on doctrine, and more on tracing the formal and informal normative commitments that characterize conflicting legal regimes,[304] and then ensuring that our institutions are capable of accommodating them. The central task is to design legal institutions with space to accommodate multiple doctrinal regimes, accounting for legal, social, and economic systems and overlapping forms of public and private power. The suggestion, above, that judges adopt an "anti-switching" presumption[305] is one illustration of how such an approach might work in existing institutions, within the judiciary. The broader challenge is how to integrate a change such as this in a way that properly situates the courts alongside other branches of government. This approach is complementary to scholarship that proposes tailored hybrid policy interventions for specific doctrinal contexts, such as copyright and antitrust law.[306] Such institutional design and targeted intervention warrant careful future consideration, drawing from an extraordinarily rich literature on legal hybridity, legal pluralism, and polycentric governance regimes.[307]

---

[302] See discussion *supra* Part III.A.

[303] *See, e.g.*, Paul Schiff Berman, *Global Legal Pluralism*, 80 S. CAL. L. REV. 1155, 1158 (2007) (citing Sally Falk Moore, *Law and Social Change: The Semi-Autonomous Social Field as an Appropriate Subject of Study*, 7 L. & SOC'Y REV. 719, 720 (1973)).

[304] *See, e.g.*, Berman, *supra* note 303, at 1157 (citing Robert M. Cover, *The Supreme Court, 1982 Term—Foreword: Nomos and Narrative*, 97 HARV. L. REV. 4, 4 (1983)).

[305] See discussion *supra* Part IV.B.1.

[306] *See generally* Noti-Victor & Tang, *supra* note 21 (discussing "targeted hybrid copyright-antitrust regulation" and arguing that it "deserves a place amongst the panoply of options").

[307] The literature here is voluminous. On legal pluralism, see, for example, John Griffiths, *What is Legal Pluralism?*, 24 J. LEGAL PLURALISM & UNOFFICIAL L. 1, 2, 5-6 (1986) (defining "'legal pluralism' as that state of affairs, for any social field, in which behavior pursuant to more than one legal occurs," then further describing concept); THE OXFORD HANDBOOK OF GLOBAL LEGAL PLURALISM (Paul Berman, ed. 2020). On polycentric governance, see the work of the Ostrom Workshop, accessible at https://ostromworkshop.indiana.edu/courses-teaching/teaching-tools/polycentric-goverance/index.html. *See also* Julia Black, *Constructing and Contesting Legitimacy and Accountability in Polycentric Regulatory Regimes*, 2 Reg. & Gov. 137 (2008) (analyzing the challenges posed by "polycentric regulatory regimes . . . in which the state is not the sole locus of authority, or indeed in which it plays no role at all").

From forging new civil society coalitions, to adapting conflict of laws principles, to adjusting the legal structures that make one regime likely to collapse into another regime, to importing lessons from legal pluralism, and beyond, we have promising resources to govern inter-regime doctrinal collapse. It is imperative that we use them.

Conclusion

In the AI context, inter-regime doctrinal collapse threatens rule of law values and empowers actors who are already best positioned to exploit overlapping, yet doctrinally and normatively distinct, regimes. This Article identifies this phenomenon and provides missing conceptual tools to better understand the tensions, contradictions, and complexities that emerge when two domains of law, like copyright law and information privacy law, apply to the same regulatory object, like data. Business as usual looks neutral. But increasingly blurry legal categories mask political choices and fortify private power. Our system of law can and should do better.

This Article's framework clarifies how and why data governance breaks down and reveals that doctrinal instability is an active site of legal and social contestation. Technology isn't inevitable, nor is the law itself. We can choose how to structure our legal doctrines. To avoid the harms that flow from unmanaged collapse of law, we need a law of collapse. This Article empowers us to create one.