Harvard Journal of Law & Technology
Volume 37, Number 3 Symposium 2023

FORMS OF DISCLOSURE: THE PATH TO AUTOMATING CLOSED BOOK PRIVACY AUDITS

Mihailis E. Diamantis, Maaz Bin Musa,***
*Lucas Ausberger*** & Rishab Nithyanand*****

ABSTRACT

The weakest link in privacy enforcement is detection. For years, agencies and activists sounded the alarm about unregulated, opaque mechanisms that organizations employ to harvest, process, and sell user data. Some state legislatures have responded in recent years by passing legislation to protect privacy rights. Federal legislation may not be far off. But privacy rights are meaningless without effective enforcement, and enforcement is blind without detection.

New techniques for uncovering privacy violations hold promise. Historically, this would have required access to data brokers' books. Unsurprisingly, such access was not forthcoming. Researchers now have tools that can carry out what this Essay calls "closed book privacy audits," detecting privacy violations without targets' cooperation. For example, closed book privacy audits can track corporate use (and misuse) of personal information across the data ecosystem by selectively feeding fictitious personal data to online platforms and measuring the impact on web experience. Automated closed book privacy audits could uncork the detection bottleneck, empowering private and public enforcers.

There is one hitch. Privacy audits require both data to test and benchmarks against which to test it. Crisp evaluative benchmarks have remained elusive. Emerging privacy laws require corporations to disclose how they collect and use personal information, but the laws do not mandate any particular form of disclosure. Through an original empirical study of privacy disclosures by California data brokers, this Essay documents the result: a widely variable mishmash of opaque representations that are impossible to audit using a consistent procedure. We argue that the law should mandate uniform privacy disclosures in a machine-readable format. Regulators could borrow from standardized disclosure frameworks used by other regulatory bodies (e.g., the United States Securities and Exchange Commission) to

* Professor of Law, University of Iowa.
** Ph.D. Candidate in Computer Science, University of Iowa.
*** M.A. Candidate in Computer Science, University of Iowa.
**** Assistant Professor of Computer Science, University of Iowa.

simultaneously improve disclosure clarity and facilitate low-cost detection of violations through closed book audits.

TABLE OF CONTENTS

"YOU CAN'T HIT WHAT YOU CAN'T SEE."[1]

I. WHY IS DATA PRIVACY COMPLIANCE STILL SO ABYSMAL?

The American public has long suspected that big tech does not exactly play fair when it comes to personal data. The Facebooks and Amazons of the world held our hands from the beginning and comfortingly assured us that they respected our privacy. They made innumerable vague commitments about what information they collected and how they used it.[2] And yet, far too often the digital world seemed to

1. Phrase used in the 1910s to describe the sidearm pitch of Washington Senators baseball great Walter Perry Johnson. HENRY W. THOMAS, WALTER JOHNSON: BASEBALL'S BIG TRAIN 387 (1998).

2. *See, e.g.*, *Privacy and Terms*, GOOGLE (July 1, 2023), https://policies.google.com/privacy?hl=en [https://perma.cc/A4ES-BGLN] ("When you use our services, you're trusting us with your information. We understand this is a big responsibility and work hard to protect your information and put you in control.").

anticipate us, offering dog treats when we adopted our first puppies and arch supports when our plantar fasciitis flared up.[3] In isolation, these incidents could be dismissed as tricks of the mind, of interest only to tin-hat theorists. Occasional high-profile incidents gave a peek at what lay behind eerie coincidence and digital déjà vu.[4] But attention spans are short, and big tech assured us that the mishaps were isolated, attributable to breaches of protocol and rogue actors.[5] Who could prove otherwise, especially with big tech refusing to let anyone look under the hood — gesturing vaguely about needing to protect "business secrets" and (ironically) privacy interests?[6]

The last few years have brought two important developments — one legal and one technological — that, if joined as this Essay proposes, could finally turn privacy rights into meaningful constraints on big tech. The legal development is the turn away from confidence in the cleansing power of notice and choice.[7] Under the old thinking, we just needed to ensure that consumers were empowered to make informed decisions. Big tech put consumers on formal notice about data practices, usually in long, technical, and loophole-riddled documents. Eventually, though, it became apparent that "notice and choice" was synonymous with "anything goes." Critics questioned whether people could meaningfully consent to what they had not read, or (if they had read) could not understand, or (if they understood) could not refuse.[8] Following the

---

3. *See generally* Lukasz Olejnik, Tran Minh-Dung & Claude Castellucia, *Selling Off Privacy at Auction*, HAL OPEN SCI. 2–3, Dec. 2013, https://inria.hal.science/hal-00915249 [https://perma.cc/M5E2-JR49].

4. *See, e.g.*, Nicholas Confessore, *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*, N.Y. TIMES (Apr. 4, 2018), https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html [https://perma.cc/4DGM-JNQQ] ("Revelations that digital consultants to the Trump campaign misused the data of millions of Facebook users set off a furor on both sides of the Atlantic.").

5. *See, e.g.*, Anita Balakrishnan, *Facebook: 'The Entire Company Is Outraged We Were Deceived' by Cambridge Analytica*, CNBC (Mar. 20, 2018), https://www.cnbc.com/2018/03/20/facebook-statement-on-cambridge-analytica-allegations.html [https://perma.cc/6E7M-PXZM].

6. *See* Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 118–20 (2019).

7. The Federal Trade Commission ("FTC") originally adopted a notice-and-choice evaluative framework. FED. TRADE COMM'N, PRIVACY ONLINE: A REPORT TO CONGRESS ii (1998), https://www.ftc.gov/sites/default/files/documents/public_events/exploring-privacy-roundtable-series/priv-23a.pdf [https://perma.cc/Q2BW-AHTG]; *see* FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE vii–viii (2012), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf [https://perma.cc/H2M7-WLXS]. The notice-and-choice framework derives from an understanding of privacy as amounting to a data subject's control over his or her information. STEPHEN J. SCHULHOFER, MORE ESSENTIAL THAN EVER: THE FOURTH AMENDMENT IN THE TWENTY-FIRST CENTURY 8–9 (2012).

8. *See generally* Kirstin Martin, *Understanding Privacy Online: Development of a Social Contract Approach to Privacy*, 137 J. BUS. ETHICS 551, 558, 561 (2015); Julia Angwin,

lead of European data regulators, some states have started to pass general data privacy laws — like the California Consumer Privacy Act ("CCPA")[9] — that set a floor for what big tech has to disclose and what companies can do with consumer data even with consent. While there is still no general privacy law at the federal level,[10] regulators have used enforcement policy to set increasingly definite standards for acceptable conduct.[11] These legal standards have made it possible to critique big tech by pointing to something other than our vague sense of unease.

As laws have delivered data protection benchmarks, computer scientists at universities in the United States and Europe have pushed forward a second development: a growing suite of technological tools for detecting corporate privacy violations. Importantly, these tools do not depend on any help from big tech, because big tech has been predictably uncooperative.[12] Some infractions are relatively straightforward to check in individual cases, like CCPA's requirement that covered websites include a "Do Not Sell My Personal Information" ("DNSMPI") button.[13] One just needs to navigate to the website and look for the button. Other infractions, such as illicit transfers of personal data, occur behind the scenes, on nonpublic servers, and through secret commercial transactions.[14] In these cases, computer scientists have learned to use what they can observe to infer what they cannot, e.g., by feeding fictionalized personal information into websites and measuring subsequent changes to a hypothetical user's web browsing experience.[15] Techniques now exist that can demonstrate data use violations with

---

Dragnet Nation: A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance 12 COLO. TECH. L.J. 291, 292 (2014); Thomas D. Haley, *Illusory Privacy*, 98 IND. L.J. 75, 122 (2022) (discussing "the futility of trying to correct the failings of platform terms in service of prolonging the notice-and-consent paradigm of privacy protection").

9. California Consumer Privacy Act of 2018, CAL. CIV. CODE § 1798.100 et seq. (West 2018).

10. *See, e.g.*, American Data Privacy and Protection Act, H.R. 8152, 117th Cong. (2022).

11. *See, e.g.*, FED. TRADE COMM'N, BRINGING DARK PATTERNS TO LIGHT 1 (2022), https://www.ftc.gov/system/files/ftc_gov/pdf/P214800%20Dark%20Patterns%20Report%209.14.2022%20-%20FINAL.pdf [https://perma.cc/5LBQ-U4FY] (defining and taking action against dark patterns).

12. Another approach to firms' unwillingness to cooperate would be to mandate public access to some subset of data. *See* Brett Frischmann & Paul Ohm, *Governance Seams*, 37 HARV. J.L. & TECH. 1115 (2024).

13. CAL. CIV. CODE § 1798.135(c)(2).

14. *See, e.g.*, Lesley Fair, *Privacy App Broke Its Privacy Promises by Disclosing Intimate Details About Users*, FED. TRADE COMM'N BUS. BLOG (Jan. 13, 2021), https://www.ftc.gov/business-guidance/blog/2021/01/health-app-broke-its-privacy-promises-disclosing-intimate-details-about-users [https://perma.cc/SQ34-8P4Y].

15. John Cook, Rishab Nithyanand & Zubair Shafiq, *Inferring Tracker-Advertiser Relationships in the Online Advertising Ecosystem Using Header Bidding*, PROC. ON PRIV. ENHANCING TECHS., Jan. 2020, at 65, 65–67.

levels of confidence that far exceed what any court would require.[16] Unlike typical audits (e.g., for food safety violations or accounting fraud) these new privacy compliance audits require no access to business records or operations. They are, so to speak, "closed book audits."

Closed book audits conducted thus far reveal a digital landscape that is distressingly unconcerned with privacy. Consider the straightforward consumer protection measure mentioned above: California's requirement that websites have a DNSMPI button. The requirement is so easy to satisfy, and violations are so patently obvious that one would expect high rates of compliance. But, according to the best available measurements, only 2% of covered websites have the button.[17] True rates of compliance are probably *much* lower because the researchers looked just for the presence of the button; they did not check whether the button worked.[18] In light of such abysmal compliance with a standard for which violations are so conspicuously visible, what hope is there for standards that big tech can violate invisibly? The closed book audits justify pessimism, having uncovered smart speakers that listen,[19] services that track children,[20] and ecosystems that broadly disseminate user data.[21]

Why does big tech act with such impunity? Once there is a law on the books, securing compliance is a matter of achieving general deterrence. General deterrence works by increasing corporations' expected costs from violating the law, thereby reducing the expected benefits of shirking. There are four steps to achieving general deterrence: detection, enforcement, punishment, and publicity. Failure at any step could explain big tech's impunity. If penalties are too small, big tech will find it more beneficial to break the law and pay later. If enforcement actions are not publicized, big tech will not know to include the risk of sanction in its business calculus.

---

16. While civil cases require only preponderance of the evidence, Conservatorship of Wendland, 28 P.3d 151, 169 (Cal. 2001) ("The default standard of proof in civil cases [in California] is the preponderance of the evidence."), closed book audits can demonstrate violations with much higher levels of statistical confidence. *See* Maaz Bin Musa & Rishab Nithyanand, *ATOM: Ad-Network Tomography*, PROC. ON PRIV. ENHANCING TECHS., July 2022, at 14 tbl.3, https://arxiv.org/pdf/2207.10791.pdf [https://perma.cc/V752-4VGH].

17. Maggie Van Nortwick & Christo Wilson, *Setting the Bar Low: Are Websites Complying with the Minimum Requirements of the CCPA?*, PROC. ON PRIV. ENHANCING TECHS., Jan. 2022, at 608, 621.

18. *Id.* at 612 ("[A]t a high-level [*sic*], on each webpage our crawler extracted the text from each hyperlink and searched it for key phrases.").

19. *See* UMAR IQBAL, POUNEH NIKKHAH BAHRAMI, RAHMADI TRIMANANDA, HAO CUI, ALEXANDER GAMERO-GARRIDO, DANIEL DUBOIS ET AL., YOUR ECHOES ARE HEARD: TRACKING, PROFILING, AND AD TARGETING IN THE AMAZON SMART SPEAKER ECOSYSTEM 2 (2023), https://arxiv.org/abs/2204.10920 [https://perma.cc/QD66-V5WP].

20. *See* Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez et al., *"Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale*, PROC. ON PRIV. ENHANCING TECHS., June 2018, at 63, 63–64.

21. *See* Bin Musa et al., *supra* note 16, at 14–15.

The first step of general deterrence — detection — remains the most persistent vulnerability. The Federal Trade Commission ("FTC") and state attorneys general have shown an increased appetite for pursuing high-profile privacy cases, imposing ever greater penalties, and publicizing their work.[22] But detecting privacy violations among the zettabytes of data that Americans generate every year still often relies on relatively old-school, labor-intensive methods like looking for "surges" in consumer complaints.[23] Consumer concern is an unreliable dowsing rod, especially for violations that are hidden from view. Studies suggest that detection is also the most crucial step for general deterrence: an increased risk of detection motivates more than does an increased penalty.[24] Neither consumers nor regulators can hit what they cannot see. Big tech learned long ago to make its data practices as unobtrusive as possible.

This Essay considers what it would take to turbocharge the first step of the enforcement pipeline. Closed book audits of big tech's data practices can provide reliable information about violations. Motivated privacy activists and researchers are already doing a lot of work.[25] As

---

22. Press Release, Fed. Trade Comm'n, Fortnite Video Game Maker Epic Games to Pay More Than Half a Billion Dollars over FTC Allegations of Privacy Violations and Unwanted Charges, (Dec. 19, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/12/fortnite-video-game-maker-epic-games-pay-more-half-billion-dollars-over-ftc-allegations [https://perma.cc/EW52-TJKE] ("The FTC's action against Epic involves two separate record-breaking settlements."); Press Release, Fed. Trade Comm'n, FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook (July 24, 2019), https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook [https://perma.cc/E6LF-4X4B] ("Facebook, Inc. will pay a record-breaking $5 billion penalty."); Press Release, Fed. Trade Comm'n, Google Will Pay $22.5 Million to Settle FTC Charges it Misrepresented Privacy Assurances to Users of Apple's Safari Internet Browser (Aug. 9, 2012), https://www.ftc.gov/news-events/news/press-releases/2012/08/google-will-pay-225-million-settle-ftc-charges-it-misrepresented-privacy-assurances-users-apples [https://perma.cc/3TGB-TBC4] ("The record setting penalty in this matter sends a clear message to all companies under an FTC privacy order. No matter how big or small, all companies must abide by FTC orders against them and keep their privacy promises to consumers, or they will end up paying many times what it would have cost to comply in the first place.") (quoting FTC Chair Jon Leibowitz).

23. Samuel Levine, Director, Consumer Protection Bureau, Fed. Trade Comm'n, Comments at University of Iowa, College of Law (Feb. 1, 2023); *see Consumer Sentinel Network*, FED. TRADE COMM'N, https://www.ftc.gov/enforcement/consumer-sentinel-network [https://perma.cc/2HET-SFXN].

24. *See* Daniel S. Nagin, *Deterrence in the Twenty-First Century*, *in* 42 CRIME AND JUST. IN AMERICA 1975–2025, at 199–202 (Michael Tonry ed., 2013) (showing that probability of detection affects deterrence calculus more than it does severity of sanction); *see also* A. Mitchell Polinsky & Steven Shavell, *On the Disutility and Discounting of Imprisonment and the Theory of Deterrence*, 28 J. LEGAL STUD. 1, 12 (1999) ("[White-collar criminals] are likely to be risk preferring in imprisonment, which suggests that less-than-maximal sanctions, combined with relatively high probabilities of apprehension, may be optimal.").

25. Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson & Christo Wilson, *Tracing Information Flows Between Ad Exchanges Using Retargeted Ads*, PROC. USENIX SEC., Aug. 2016, at 481, 481–82, https://www.usenix.org/system/files/conference/usenix

other scholars have proposed, public enforcers should be in regular contact (if not partnership) with these outside parties.[26] However, closed book audits remain resource intensive, limiting how much activists and researchers can achieve.

Below, we describe measures that would streamline the detection process. We propose bringing the two developments discussed above — privacy law disclosure requirements and closed book audits — into closer conversation with each other. Privacy disclosure mandates should include content and format standards that are more amenable to closed book audits. Vague, inconsistent, and varied language in privacy policies makes closed book audits costly or impossible. Improved formatting could help researchers conduct audits at scale. Machine-readable disclosures could even pave an eventual path to automated audits.

After providing an overview of present-day privacy disclosure requirements (Part II), we turn to the centerpiece of our argument, a first-of-its-kind, industry-wide examination of corporate privacy policies (Part III). Prior research has parsed policy language, but it has not tested that language against CCPA mandates.[27] We find that privacy policies are highly variable, even when comparing functionally equivalent policy sections. Their words, their readability scores, and even their lengths exhibit little uniformity. This variability frustrates closed book audits because researchers must parse the nuanced language of each individual policy, making controvertible interpretive judgments when translating the language into an auditable commitment.[28] To remedy the problem, we propose a regime of privacy disclosure mandates that

---

security16/sec16_paper_bashir.pdf [https://perma.cc/KPN5-LTGC]; Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Rivachander, Ziqi Wang, Joel Reidenberg et al., *MAPS: Scaling Privacy Compliance Analysis to a Million Apps*, PROC. ON PRIV. ENHANCING TECHS., July 2019, at 66, 66–67, https://usableprivacy.org/static/files/popets-2019-maps.pdf [https://perma.cc/N2LE-QH52].

26. *See* Nataliia Bielova, Cristiana Santos & Colin M. Gray, *Two Worlds Apart! Closing the Gap Between Regulating EU Consent and User Studies*, 37 HARV. J.L. & TECH 1293 (2024).

27. Kanthashree Mysore Sathyendra, Shomir Wilson, Florian Schaub, Sebastian Zimmec & Norman Sadeh, *Identifying the Provision of Choices in Privacy Policy Text*, CONF. ON EMPIRICAL METHODS IN NAT. LANGUAGE PROCESSING, Sept. 2017, at 2774, 2774–75, https://aclanthology.org/D17-1294.pdf [https://perma.cc/E3VL-XNJZ]; Hamza Harkous, Kassem Fawaz, Florian Schaub, Kang G. Shin & Karl Aberer, *Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning*, PROC. USENIX SEC., Feb. 2018, at 531, 531–48, https://www.usenix.org/system/files/conference/usenixsecurity18/sec18-harkous.pdf [https://perma.cc/HWH8-PYDD].

28. *See, e.g.*, *Your Echos are Heard: Tracking, Profiling, and Ad Targeting in the Amazon Smart Speaker Ecosystem*, ALEXA ECHOS, https://alexaechos.com [https://perma.cc/4HKS-NU4G] ("Amazon [publicly stated] they *'do not use voice recordings to target ads.'* While Amazon may not literally be using the 'recordings' (as opposed to transcripts and corresponding activities), our results suggest they are processing voice recordings, inferring interests, and using those interests to target ads. This distinction between voice recordings and processed recordings may not be meaningful to many users.").

would facilitate closed book audits (Part IV). Many other areas of law require corporate disclosures with well-defined formats, and privacy law could benefit from their example. As proof of concept, we offer a specific disclosure form tailored to a CCPA disclosure mandate and contrast our proposal with Platform for Privacy Preferences ("P3P"), a failed effort to standardize disclosures in Europe. Formulaic disclosures and the closed book audits they facilitate could help open the detection bottleneck that is choking effective privacy enforcement.

II. BACKGROUND TO THE CCPA DISCLOSURE REQUIREMENTS

The CCPA has been in effect since 2020. As the most comprehensive privacy law in the United States, the CCPA establishes several new rights that give consumers control over how companies collect and disperse their data. To secure these new rights, the CCPA imposes corresponding obligations on covered businesses[29] and data brokers,[30] particularly obligations to make various disclosures. Some other states including Colorado,[31] Connecticut,[32] Iowa,[33] Utah,[34] and Virginia[35] have data privacy statutes too. While different in detail, their structure broadly resembles the CCPA.[36]

The CCPA's foundation is the various rights it guarantees for consumers. These include:

(1) The right to know which categories of personal information a firm collects, the purpose of such collection, and whether (and to whom) the firm sells or discloses the information.[37]

(2) The right to delete any personal information a firm gathers.[38]

---

29. Covered businesses are for-profit entities that do business in the State of California and satisfy one or more of the following thresholds: (a) have annual gross revenues in excess of $25 million in the prior calendar year, (b) buy, sell, or share the personal information of $100,000 or more consumers or households, or (c) derive fifty percent or more of annual revenue from selling or sharing consumers' personal information. California Consumer Privacy Act of 2018, CAL. CIV. CODE § 1798.140(d) (West 2018).

30. A data broker is a business "that knowingly collects and sells to third parties the personal information of a consumer with whom the business does not have a direct relationship." *Id.* § 1798.99.80(d).

31. *Colorado Privacy Act*, COLO. REV. STAT. § 6-1-1301 (2023).

32. *Connecticut Data Privacy Act*, CONN. GEN. STAT. § 42-522 (2022).

33. *Iowa Consumer Data Protection Act*, S.F. 262 (2023).

34. *Utah Consumer Privacy Act*, UTAH. CODE. §§ 13-61-101–404 (2022).

35. *Virginia Consumer Data Protection Act*, VA. CODE § 59.1-575 (2023).

36. For a comparison of the various state privacy laws, see *US State Privacy Legislation Tracker*, INT'L ASS'N OF PRIV. PROFESSIONALS (July 7, 2023), https://iapp.org/resources/article/us-state-privacy-legislation-tracker [https://perma.cc/F9LE-T8R8].

37. CAL. CIV. CODE § 1798.100 (West 2018).

38. *Id.* § 1798.105.

(3) The right to correct inaccurate personal information that a firm holds.[39]

(4) The right to access personal information a firm gathers.[40]

(5) The right to opt out of a firm selling and sharing personal information.[41]

(6) The right to limit the use and disclosure of sensitive personal information.[42]

(7) The right against corporate retaliation for exercising CCPA consumer rights.[43]

The CCPA also imposes several corresponding obligations on covered businesses. These include maintaining an up-to-date privacy policy[44] that discloses the following:

(1) The rights consumers have under the CCPA.[45]

(2) Two or more methods by which consumers may exercise their rights.[46]

(3) The categories, sources, and purposes of any personal information a firm collects, shares, or sells.[47]

(4) The categories of any personal information disclosed to third parties.[48]

(5) The mechanisms for consumers to opt out of disclosure of personal information.[49]

Most scholars have viewed privacy disclosures as tools for empowering consumers.[50] Notice-and-consent theorists believe consumer consent cleanses otherwise suspect practices.[51] Disclosures put consumers in the driver's seat, the thinking goes, by advising them about the nature of the commercial exchange they will enter if they proceed.[52] Informed

39. *Id.* § 1798.106.
40. *Id.* § 1798.110.
41. *Id.* § 1798.120.
42. *Id.* § 1798.121.
43. CAL. CIV. CODE § 1798.125.
44. *Id.* § 1798.130(a)(5).
45. *Id.* § 1798.130(a)(5)(A).
46. *Id.*
47. *Id.* § 1798.130(a)(5)(B).
48. *Id.* § 1798.130(a)(5)(C).
49. CAL. CIV. CODE § 1798.135(c)(2).
50. *See* Robert H. Sloan & Richard Warner, *Beyond Notice and Choice: Privacy, Norms, and Consent*, 14 J. HIGH TECH. L. 370, 370–71 (2014) (outlining the usual case made on behalf of notice and choice).
51. *See id.*
52. *See id.*

consumers will, the thinking goes, steer clear of interactions that they deem harmful.[53]

The notice-and-consent model faces some powerful critiques. Critics observe that disclosures are too long (it would take over thirty workdays for the average person to read all the privacy policies they encounter each year),[54] too complex (most privacy policies require a college-level reading ability),[55] too misleading (with both legal and technical loopholes),[56] and too hard to find.[57] If consumers cannot read or understand disclosures, this form of notice puts them in no better position — there is no meaningful sense in which consumers can consent.

This Essay departs from existing legal scholarship by viewing privacy disclosures in a different light. Regardless of whether mandatory disclosures actually empower consumers, they are devices for forcing firms to take on a measure of legal vulnerability.[58] Disclosures in part define what counts as a violation. Collecting email addresses violates the CCPA only if a firm fails to disclose the practice. Disclosures establish the benchmarks. When disclosures are sufficiently definite, they can be audited using closed book methods. By getting privacy disclosures right, the law could set the stage for procedures that meaningfully verify compliance and detect breaches. However, as the next Part demonstrates, present-day privacy disclosures are riddled with obstructive ambiguity and variability.

III. MEASURING THE MANY MEANINGS OF DISCLOSURE

The CCPA mandates specific disclosures but stops short of mandating any disclosure format. Firms might take two approaches with this freedom. They might naturally coalesce around similar disclosure templates as industry standards emerge, or each firm might blaze its own trail. This latter possibility is concerning. It would open space for firms to undermine the consumer-facing function of disclosures. For example, firms could employ idiosyncratic dark patterns in the text of

---

53. *See id.*

54. Aleecia M. McDonald & Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*, 4 J.L. & POL'Y FOR INFO. SOC'Y 543, 563 (2008).

55. Patrick Gage Kelly, Lucian Cesca, Joanna Bresee & Lorrie Faith Cranor, *Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach*, PROC. SIGCHI CONF. ON HUMAN FACTORS IN COMP. SYS. 1573, 1573 (2010).

56. Alex Kozinski & Mihailis E. Diamantis, *An Eerie Feeling of Déjà Vu: From Soviet Snitches to Angry Birds*, *in* CAMBRIDGE HANDBOOK OF SURVEILLANCE LAW 420, 425–26 (David Gray & Stephen E. Henderson eds., 2017); *see generally* Irene Pollach, *What's Wrong with Online Privacy Policies?*, 50 COMM. ACM 103 (2007).

57. Lori Andrews, *A New Privacy Paradigm in the Age of Apps*, 53 WAKE FOREST L. REV. 421, 463 (2018) ("[P]rivacy policies are routinely difficult to find.").

58. James C. Cooper, *Does Privacy Want to Unravel?*, 37 HARV. J.L. & TECH. 1037 (2024) (arguing that one role for regulators is to help firms credibly commit to privacy policies).

their privacy policies that prevent consumers from understanding what happens to their data.[59] More relevant to the concerns of this Essay, unpredictable formats also undermine disclosures' verification function by impeding large-scale compliance audits.

We undertook an original study of CCPA disclosures to see which path firms have taken in California. The goal was to learn how closely firm disclosures resemble each other in the absence of a mandatory format. Using a novel method for automatic labeling and clustering of CCPA disclosures within privacy policies, we uncovered a widely disparate patchwork of approaches. This variability poses a challenge for any large-scale or automated effort to assess compliance with the CCPA.

Our findings highlight the best-case scenario that consumers and regulators face regarding privacy disclosures. The study departs from existing work on privacy policies by focusing on data brokers, rather than apps and websites. Data brokers are the central parties through which all data (including data collected by websites and apps) flows into the information economy.[60] Though "data brokers are presently subject to very little federal or state oversight," both the FTC and the Consumer Financial Protection Bureau have signaled their intent to change that.[61] The CCPA has no substantive provisions specific to data brokers, but it does require them to register their name, physical address, email address, and website URL.[62] Because data broker information is publicly available on a website maintained by the California Attorney General, generating a corpus of data broker privacy policies is relatively straightforward.[63] California-registered data brokers are appealing subjects for three reasons: (1) data brokers are reliably more sophisticated than the hobbyists who design many apps and websites, (2) many data brokers are obligated under California law to publish privacy policies,[64] and (3) the act of registration reflects awareness of the disclosure obligation. Consequently, data broker privacy policies should set the bar for what we can expect of privacy disclosures in the absence of a mandatory format.

---

59. Colin M. Gray, Nataliia Bielova, Michael Toth, Cristiana Santos & Damian Clifford, *Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective*, CHI. CONF. ON HUM. FACTORS IN COMPUTING SYS., May 2021, at 1, 2.

60. *What Is a Data Broker?*, MCAFEE (Apr. 4, 2022), https://www.mcafee.com/blogs/tips-tricks/what-is-a-data-broker [https://perma.cc/N45J-7LFA].

61. Kirk Nahra, Ali Jessani & Samuel Kane, *CFPB Issues Request for Information on Data Brokers*, COMPLIANCE & ENFORCEMENT (Apr. 10, 2023), https://wp.nyu.edu/compliance_enforcement/2023/04/10/cfpb-issues-request-for-information-on-data-brokers [https://perma.cc/ANE6-PZ8Y].

62. California Consumer Privacy Act of 2018, CAL. CIV. CODE § 1798.99.82(b)(2)(A) (West 2018).

63. *Id.* § 1798.99.84.

64. *Id.* § 1798.100(a).

Readers wishing to skip the technical details of our experimental design should jump to Section B, which reports our results.

A. Experimental Design

Our goal was to develop an automated framework for identifying how data broker privacy policies represent each required CCPA disclosure. At a high level, we achieved this by: (1) gathering a corpus of privacy policies; (2) manually annotating a subset of them (the ground truth dataset); (3) training and validating a language classifier with the ground truth dataset; (4) using the classifier to segment the policies into contextually coherent chunks of text; (5) developing a machine learning classifier to group chunks of text that are responsive to specific CCPA disclosure mandates; and (6) clustering chunks within groups to identify common patterns in disclosure representations. An overview of this process is illustrated in Figure 1: Overview of Methodology for Identifying Common Disclosure Patterns.

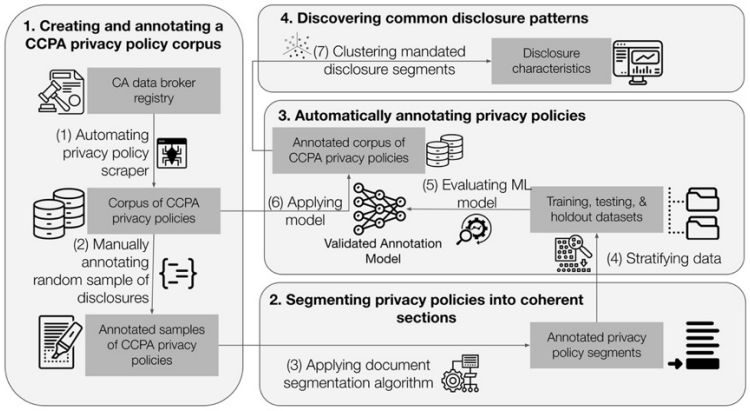


Figure 1: Overview of Methodology for Identifying Common Disclosure Patterns

1. Creating and Annotating a Privacy Policy Corpus

The first step in implementing our study was to gather our corpus of privacy policies. As noted above, the CCPA applies only to businesses that satisfy data or revenue thresholds.[65] Prior work has shown that identifying the businesses that are subject to the CCPA is itself a challenging task due to the general unavailability of data about whether

65. *Id.* § 1798.140(d).

businesses satisfy the CCPA's criteria.[66] To circumvent this challenge, we obtain privacy policies from data brokers that have self-identified as being subject to the CCPA and registered with the California Attorney General. We created our privacy policy corpus by scraping the "Data Broker Registry" for the websites associated with registered data brokers.[67] At the time of our data gathering (November 2022), 468 data brokers were registered and selected for this study.

We then used an automated web crawler to visit the websites associated with each registered data broker and searched their front pages for a link to their privacy policy. Of the 468 registered data brokers, 426 (ninety-one percent) had a working website address and 326 (seventy percent) had a privacy policy that mentioned the keywords "CCPA" or "California." These 326 privacy policies formed our corpus.

To develop and validate our automated techniques, we randomly sampled one hundred (thirty-one percent) privacy policies in the corpus to create a ground truth dataset. We manually annotated sections of text that were associated with specific CCPA-mandated disclosures for each sampled policy. Sections were annotated with one or more of the labels listed in Table 1: Annotation Labels Assigned to Sentences Within Privacy. Two of the authors of this Essay were responsible for manually annotating each of the sampled policies.

Table 1: Annotation Labels Assigned to Sentences Within Privacy

Label	**Description**
Methods	Designated methods for exercising CCPA-granted rights.
Update	Date of last policy update.
Description	Description of CCPA-granted consumer privacy rights.
PII-Collected	Categories of personally identifiable information collected.
PII-Sold	Categories of personally identifiable information sold or disclosed.
Other	Text not related to any of the above descriptions (including non-CCPA text).

66. *See* Nortwick et al., *supra* note 17, at 622.

67. *Data Broker Registry*, CAL. DEP'T JUST., OFF. ATT'Y GEN., https://oag.ca.gov/data-brokers [https://perma.cc/4V8P-4KD3].

2. Segmenting Privacy Policies into Coherent Sections

During manual annotation of the ground truth dataset, we realized that even when privacy policies completely represent a firm's data collection practices, they are often poorly formatted. Most policies did not include structural elements, such as paragraphs and section headings, that would assist in easily parsing and understanding the document. It was often unclear which sections of the policy were associated with which requirements of the CCPA. Such structural breaks are a prerequisite for any automated auditing system. Therefore, we applied a variety of computational linguistic approaches (including text tiling[68] and graph-based segmentation[69]) for automatically segmenting privacy policies into contextually coherent chunks of text — i.e., groups of contiguous sentences that are semantically and topically related.

3. Automatically Annotating Privacy Policy Segments

Our next goal was to automatically annotate each of the segments generated from the prior step with one or more of the annotation labels described in Table 1: Annotation Labels Assigned to Sentences Within Privacy. To accomplish this task, we created one machine learning classification model for each annotation label. To start, we randomly selected ten percent of the segments associated with each annotation label to serve as a "holdout" validation dataset that we did not use in training the models. We used the remaining ninety percent of annotated segments to train our models. After combining the individual classifiers, we had an ensemble model that used all our labels to classify policy segments.[70]

We evaluated our ensemble model using our holdout validation dataset. Table 2 shows the precision, recall, and most common error associated with each classifier in our ensemble. A higher precision for a specific label indicates that the classifier is less likely to misapply the label to an unrelated segment. A higher recall for a label indicates that the classifier is less likely to misclassify segments associated with that label. To put our results in context, prior work shows that labeling

68. *See generally* Marti A. Hearst, *TextTiling: Segmenting Text into Multi-Paragraph Subtopic Passages*, 23 COMPUTATIONAL LINGUISTICS 33 (1997).

69. *See generally* Goran Glavaš, Federico Nanni & Simone Paolo Ponzetto, *Unsupervised Text Segmentation Using Semantic Related Graphs*, PROC. FIFTH JOINT CONF. LEXICAL & COMPUTATIONAL SEMANTICS, Aug. 2016, at 125.

70. An ensemble model combines multiple models to achieve better predictive performance than the individual models. *See generally* Josef Kittler, Mohamad Hatef, Robert P.W. Duin & Jiri Matas, *On Combining Classifiers*, 20 IEEE TRANSACTIONS ON PATTERN ANALYSIS & MACH. INTEL. 226, 238 (1998).

privacy policy language is challenging even for trained human judges.[71] We cannot expect a machine to classify segments better than trained experts.

Table 2: Performance of Our Automated Segment Annotator on Various Segment Types

Segment Type	Precision	Recall	Most Frequent Incorrect Label
Methods	74.4%	73.6%	Description
Update	62.2%	79.6%	Description
Description	90.3%	84.0%	PII-Collected
PII-Collected	88.7%	94.4%	Other (non-CCPA)
PII-Sold	75.0%	48.0%	PII-Collected

To better understand the limitations of our classifier, we also analyzed the annotation labels that were most frequently mis-associated with each segment type. This highlighted a limitation of our classifier: it was generally incapable of reliably distinguishing between descriptions of data collection and descriptions of data disclosures (i.e., it often confused "PII-Sold" segments for "PII-Collected" segments). Because of our classifier's poor performance in identifying "PII-Sold" segments, we do not present conclusions from our analyses of segments associated with the "PII-Sold" annotation type. Considering all other annotation types, our results show that the ensemble performed reasonably well at distinguishing between CCPA and non-CCPA segments in policies and between the various CCPA-mandated disclosures.

4. Discovering Common Disclosure Representation Patterns

To this point, we had collected our corpus of privacy policies, broken them into topical segments, and grouped segments that were responsive to specific CCPA disclosure mandates. We then looked for common patterns within each group using an automated machine learning-based text clustering approach. Automated clustering approaches generally measure the similarity between two input samples and group similar input samples within a "cluster" (in our case, privacy policy text segments that are responsive to the same CCPA disclosure mandate).

71. Trained human judges tasked with annotating a variety of disclosures made in privacy policies only achieved annotation agreement rates of 88–98%. *See* Anthony D. Miyazaki & Sandeep Krishnamurthy, *Internet Seals of Approval: Effects on Online Privacy Policies and Consumer Perceptions*, 36 J. CONSUMER AFF. 28, 36–38 (2002), https://www.jstor.org/stable/23860158 [https://perma.cc/3FBS-YLQN].

After experimenting with a range of different methods, we settled on one that provided the highest quality clusters. Our chosen method started with a standard technique called "Latent Dirichlet Allocation" ("LDA"). LDA analyzes the frequency of words across a corpus of text to identify a number of "topics," i.e., groups of words that are likely to be related to each other.[72] It allowed us to represent each document in the corpus as a mixture of topics. We used K-means clustering to cluster documents into "K" groups based on the similarity of their topic mixtures. To determine the optimal number of clusters, we used a silhouette analysis.[73] Table 3 shows the silhouette scores that our approach returned for each segment type. Silhouette scores range from +1 to -1, with a score close to +1 indicating higher quality clusters of more similar segments.

Table 3: Silhouette Scores and Configurations for Best Performing Clustering Model for Each Segment Type[74]

Segment Type	**LDA Topics**	**K-means Clusters**	**Silhouette Score**
Methods	3	3	.68
Update	6	7	.59
Description	11	12	.52
PII-Collected	6	6	.55

B. Results: Disclosures Are Varied and Vague

We used the method described above to automatically annotate disclosures in data broker privacy policies; group segments that responded to the same CCPA disclosure mandate; and compare the similarity of segments within each group. We assessed each group of disclosures to determine: (1) the rate of data brokers' compliance with CCPA disclosure mandates; (2) the readability characteristics of disclosures; and (3) the variability of language used in disclosures.

---

72. *See generally* David M. Blei, Andrew Y. Ng & Michael I. Jordan, *Latent Dirichlet Allocation*, 3 J. MACH. LEARNING RSCH. 993 (2003).

73. *See generally* Peter J. Rousseeuw, *Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis*, 20 J. COMPUTATIONAL & APPLIED MATHEMATICS 53 (1987). The silhouette score is a metric to evaluate the cohesion and separation of clusters.

74. We do not include "PII-Sold" in our analysis because of our automated annotator's poor performance in identifying related segments.

1. Compliance with CCPA Disclosure Mandates

Table 4 paints a promising picture of data broker compliance with CCPA disclosure mandates. Between 94.8% and 98.7% of all policies in our corpus had at least some text segments related to each disclosure mandate. This suggests California data brokers that have privacy policies that explicitly reference the CCPA are generally compliant with CCPA disclosure requirements. Our manual inspection of a sample of policies labeled as "non-compliant" by our annotator showed that despite the occurrence of a few incorrectly labeled segments within policies, most were correctly identified as non-compliant due to missing disclosures.

Table 4: Fraction of Surveyed Policies Containing Segments Annotated with the Corresponding Label

Annotation Label	**Policies with Related Segments**
Methods	94.8%
Update	95.2%
Description	98.7%
PII-Collected	97.0%

2. Readability of CCPA-Mandated Disclosures

Current CCPA disclosure mandates are intended to improve consumers' understanding of and control over the use of their data.[75] Therefore, it is important that the policies are comprehensible to average users. To be comprehensible, a policy must: (1) have an intuitive structure so that readers can easily identify specific disclosures; (2) use simple language; and (3) be a reasonable length. Incidentally, these same three features would also help automated auditing systems to extract information regarding businesses' data handling practices.

Our study demonstrates that privacy policies are poorly and inconsistently structured. Privacy policies with paragraph headings or intuitive section breaks would have been easy to parse and compare. However, we needed to develop and train a sophisticated policy segmentation algorithm just to recognize when policies moved from

75. California Privacy Rights Act of 2020, Proposition 24, § 2G ("The State therefore has an interest in mandating laws that will allow consumers to understand more fully how their information is being used, and for what purposes . . . . Additionally, if a consumer can tell a business not to sell the consumer's data, then that consumer will not have to scour a privacy policy . . . .").

discussing one topic to another. This strongly suggests that current privacy policies are not reliably structured in a way that is easy for humans, let alone automated systems, to process.

As to disclosure length, we find that even seemingly simple disclosures are often excessively verbose and difficult to parse. Figure 1 (Left) shows the number of words in segments related to each disclosure mandate. The "Update" disclosure requires firms only (1) to state the date of last privacy policy update and (2) to update the policy at least every twelve months. Firms managed to complicate even this straightforward mandate. Puzzlingly, some businesses were observed to be marking the date of the last update on individual sections of the policy, rather than on the entire policy. On average, "Update" disclosure segments were 220 words long.

As to readability, CCPA disclosures require an average of eleven to seventeen years of education to understand, as measured by the Flesch-Kincaid Grade Level readability score ("FKGL"). FKGL is a U.S. Military Standard for technical documents.[76] Figure 2 (Right) breaks out the data by disclosure type. By way of comparison, several states mandate a FKGL score under nine for insurance policy documents.[77] Even the seemingly straightforward "Update" and "Methods" disclosures have higher FKGL scores.

76. J. PETER KINCAID, ROBERT P. FISHBURNE, RICHARD L. ROGERS & BRAD S. CHISSOM, DERIVATION OF NEW READABILITY FORMULAS (AUTOMATED READABILITY INDEX, FOG COUNT AND FLESCH READING EASE FORMULA) FOR NAVY ENLISTED PERSONNEL 19–20 (1975). The FKGL metric uses a combination of average sentence length and average word complexity to estimate the number of years of education required to understand a passage of text.

77. *See, e.g.*, *Adoption of Flesch Reading Ease Test*, TEX. DEP'T INS. (June 15, 1992), https://www.tdi.texas.gov/pubs/pc/pccpfaq.html [https://perma.cc/5KHJ-6H5R]; 3 COLO. CODE REGS. § 702-5-1-18-6 (2023).

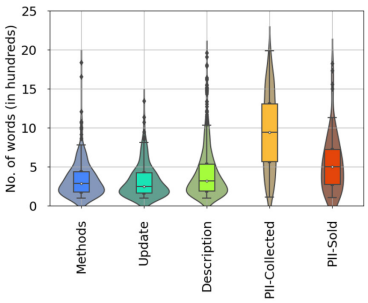


Figure 1 (Left): Distribution of Text Segment Word Counts (in Hundreds) for Each Segment Type

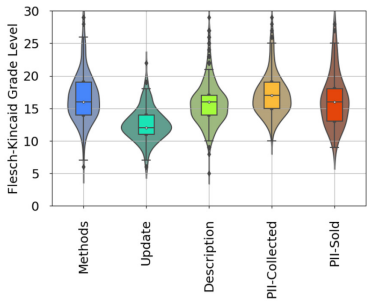


Figure 2 (Right): Distribution of Text Segment Readability Metrics (Flesch-Kincaid Grade Level) for Each Segment Type

3. Variety of Disclosure Styles and Representations

To assess the variability of language used in privacy disclosures, we focused on the "Methods" mandate. This mandate essentially requires that businesses disclose two methods of contact (e.g., an email address and a phone number) for consumers to exercise their CCPA-granted rights. We select this specific mandate for our analysis for three reasons: (1) the possibility of complying with the mandate in a succinct manner; (2) the smaller word counts in actual disclosures related to the mandate (median of 240 words); and (3) the smaller number of representational styles identified in our cluster analysis (just three clusters). Together, these considerations suggest that the variations observed in this specific disclosure will serve as a lower bound for the variability in other, more complicated disclosure mandates.

Figure 3 illustrates how the three LDA-derived topics were distributed across the K-clusters of "Methods" disclosure text. Each dot in the figure represents a "Methods" disclosure in a policy from the corpus. Each axis in the three-dimensional plot corresponds to one of the three LDA-derived topics. The three different shapes correspond to the three K-clusters. The position of a dot in the grid indicates how much of each topic is contained within the disclosure the point represents. Points at the center of the grid contain a near equal mix of each of the three LDA topics, while points at the origin of one of the axes contain a mix of only two of the LDA topics. Our high cluster silhouette and topic coherence scores (see Table 3 above) indicate that these representations are a good approximation of the types of representations that exist in our corpus of privacy policies. Notably, most dots are scattered

between the vertices, each reflecting a different mix of elements from each cluster.

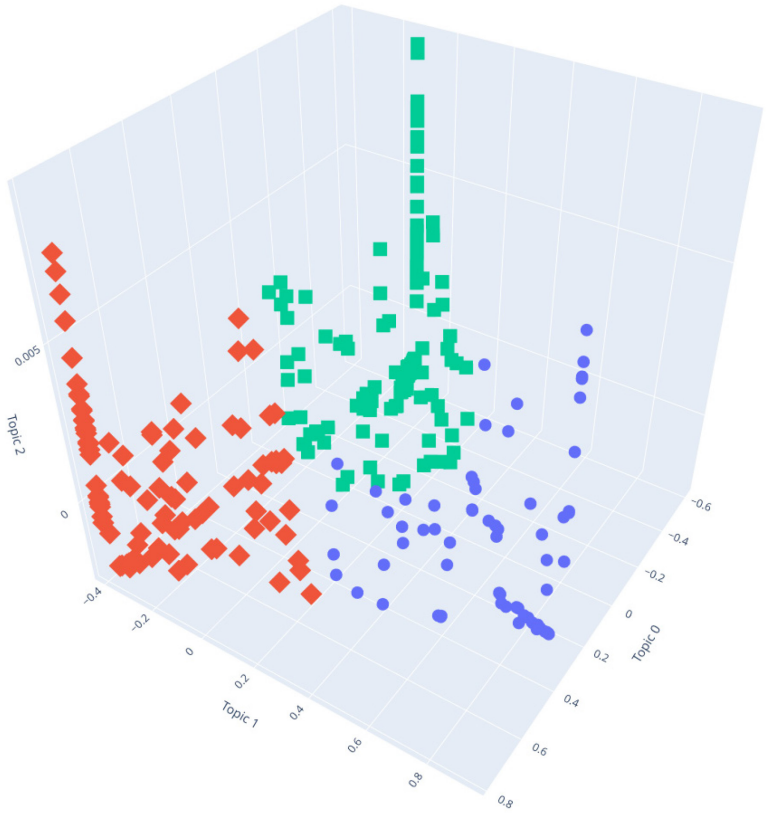


Figure 3: Three-Dimensional Representation of the Distribution of Identified Topics Within "Methods" Disclosures[78]

Our data establishes two key conclusions about the approach privacy policy disclosures may take in the absence of a prescribed format. First, even disclosures pertaining to seemingly simple mandates could be much shorter, simpler, and better structured. Second, businesses' wide variety of approaches to disclosure is unnecessarily cumbersome for both humans and automated systems to navigate.

78. Each point represents the "Methods" disclosures extracted from one privacy policy, each axis corresponds to one of our three LDA-derived topics, and each color corresponds to the cluster within which the segment was placed.

C. Discussion: An Impediment to Closed Book Audits

The approach we took in our study essentially amounts to a best-effort attempt to automatically extract CCPA-related disclosures from existing privacy policies. The hurdles we faced correspond to challenges that consumers face in exercising control over their data and that auditors must confront in holding businesses accountable. We highlight each of these challenges below.

1. Availability of Privacy Policies

Privacy policies are unnecessarily difficult to find. We focused on registered data brokers: a class of businesses that by definition is already compliant with one aspect of the CCPA (the registration requirement). By virtue of the registration process, California maintains a central database of website addresses for each data broker. Nonetheless, we had difficulty obtaining privacy policies for our analysis. We were able to identify and analyze the CCPA-specific disclosures of only seventy percent of the data brokers on the 2022 version of the data broker registry. We expect that this number represents an upper bound on compliance with the requirement to publish a privacy policy. It may be more difficult to locate the privacy policies of other covered entities that need not register with the California Attorney General.

2. Poor Document Formatting and Structure

Our human and machine annotators struggled to navigate the privacy policies in our corpus. Policies often lack an organized structure, like topic-specific paragraphs and sections. This impedes human readability and complicates the task of developing policy segmentation tools for automated audits. Though we used state-of-the-art unsupervised segmentation techniques, our policy segments still contained errors. Each of our methods generated segments that either included extraneous text not related to a specific disclosure or excluded text related to a specific disclosure. Our segmentation error rates are reasonable for demonstrative purposes, but they are bound to negatively impact any attempt to automate compliance audits.

3. High Variability in Disclosure Styles and Representations

Our results show that even simple disclosures such as the "Update" and "Methods" requirements of the CCPA exhibit a wide variability. They often appear in unexpected parts of a privacy policy or lumped into segments describing other disclosures. Once again, these practices make information extraction by humans and machines unnecessarily

cumbersome. These challenges only grow when attempting to extract information about more nuanced mandates. For example, our automated systems could not distinguish between segments related to the "PII-Collected" and "PII-Sold" mandates with any measure of confidence.

IV. STANDARDIZING PRIVACY DISCLOSURES

The CCPA requires businesses that handle personal data to make various disclosures, but the law does very little to specify the form or format the disclosures must take. Part III documents the result: businesses take widely disparate approaches that almost seem calculated to evade easy audit. This Part shows what more exacting disclosure requirements could look like and how they would facilitate closed book audits.

A. Legal Framework and Precedents

The CCPA is a relatively late entrant into the world of mandatory disclosures. Each spring, millions of taxpayers prepare to disclose detailed personal income information to the Internal Revenue Service. Manufacturers in forty-one different industrial sectors file annual reports about greenhouse gas emissions to the Environmental Protection Agency.[79] Publicly traded corporations submit quarterly financial statements and disclose market risks to the Securities and Exchange Commission ("SEC"). These legal regimes exist to help government agencies keep tabs on conduct that would otherwise be hidden from view.

Tax returns, emissions reports, and SEC filings have two features in common that CCPA disclosure mandates lack. First, these existing regimes provide forms for the disclosures, such as Form 1040 for individual tax returns, the Electronic Greenhouse Gas Reporting Tool for emissions,[80] or dozens of forms for corporate financial health.[81] Forms make disclosures predictable by standardizing what information is provided, the format it takes, and where it appears. The CCPA provides nothing analogous. Several private companies offer templates to facilitate CCPA disclosures, but California has no format mandate, standard, or recommendations.

---

79. 40 C.F.R. § 98.

80. ENV'T PROT. AGENCY, ELECTRONIC GREENHOUSE GAS REPORTING TOOL (E-GGRT) 4 (2013), https://www.epa.gov/sites/default/files/2015-07/documents/subpartmmeggrtreportingwebinar2013.pdf [https://perma.cc/3957-WV3F].

81. *Forms List*, SEC. & EXCH. COMM'N, https://www.sec.gov/forms [https://perma.cc/4J7P-DNNC].

A second major difference between CCPA disclosures and IRS, EPA, or SEC disclosures is that the latter are filed with a government body. Filing requirements raise the stakes for firms by (1) conveying that the government cares about an issue, (2) forcing filers to confirm verifiable facts, and (3) potentially entailing serious penalties for misrepresentations.[82] By contrast, the CCPA only requires firms to publish privacy disclosures on company websites; nothing needs be filed with California authorities.[83]

The SEC's disclosure model is particularly relevant for privacy enforcers because the SEC and the FTC face enforcement mandates that are similarly incommensurate to agency resources.[84] The SEC oversees annual trading of approximately $350 trillion, across twenty-four national securities exchanges, by 29,000 registered entities and 5,248 publicly traded companies.[85] This would be an impossible task for the 4,500 staff members of the SEC to monitor alone.[86] Likewise, the FTC's "consumer protection mission alone covers almost the entire economy,"[87] and its staff number under 1,100.[88]

To overcome its deficit of resources, the SEC devised a system that effectively recruits private parties to help. The SEC publishes company

---

82. 18 U.S.C. § 1001 ("[W]hoever, in any matter within the jurisdiction of the executive, legislative, or judicial branch of the Government of the United States, knowingly and willfully . . . makes any materially false, fictitious, or fraudulent statement or representation . . . shall be fined under this title, imprisoned not more than 5 years."). The FTC has used this sort of commitment-and-enforcement mechanism before, e.g., in forcing Facebook CEO Mark Zuckerberg to personally certify and submit privacy statements. FED. TRADE COMM'N, *supra* note 22 ("Facebook CEO Mark Zuckerberg and designated compliance officers must independently submit to the FTC quarterly certifications that the company is in compliance with the privacy program mandated by the order, as well as an annual certification that the company is in overall compliance with the order. Any false certification will subject them to individual civil and criminal penalties.").

83. The CCPA simply states that a covered business must include required disclosures "in its online privacy policy or policies . . . or if the business does not maintain those policies, on its internet website." California Consumer Privacy Act of 2018, CAL. CIV. CODE § 1798.130(a)(5).

84. *See About the SEC*, SEC. & EXCH. COMM'N (Nov. 22, 2016), https://www.sec.gov/about [https://perma.cc/T77J-R9LR] ("The mission of the SEC is to protect investors; maintain fair, orderly, and efficient markets; and facilitate capital formation.").

85. *About the SEC*, SEC. & EXCH. COMM'N (Apr. 6, 2023), https://www.sec.gov/strategic-plan/about [https://perma.cc/ZS6J-LAUU].

86. *Id.*; Jill E. Fisch, *Class Action Reform, Qui Tam, and the Role of the Plaintiff*, 60 L. & CONTEMP. PROBS. 167, 199 (1997) ("The Securities and Exchange Commission ('SEC') repeatedly has acknowledged, for example, that private litigation enables a level of compliance that would be impossible to achieve if enforcement were limited to the government.").

87. Samuel Levine, Director, Bur. of Consumer Prot., Fed. Trade Comm'n, Believing in the FTC, Keynote Address at the *Harvard JOLT*-UIowa IBL Symposium: Beyond the FTC (Apr. 1, 2023), https://www.ftc.gov/system/files/ftc_gov/pdf/Remarks-to-JOLT-4-1-2023.pdf [https://perma.cc/J3HY-DGLX].

88. *Federal Trade Commission*, U.S. EEOC, https://www.eeoc.gov/federal-sector/federal-trade-commission-ftc-0 [https://perma.cc/4E26-WZQK].

filings in a publicly accessible database called EDGAR.[89] Private investors then parse these forms, looking for information that will help them decide whether to purchase or sell stocks based on their prediction of a company's future performance. This sets up a mechanism through which investors and the market, in a sense, "audit" companies' SEC filings. Most of the time, a company will perform within a range that investors, relying on public disclosures, anticipated. However, if a company substantially underperforms expectations, investors suffer a loss and look for opportunities to recoup. They return to company filings, inspecting them for material misstatements or omissions that may have induced investment decisions. Perhaps, to use an example from recent memory, the company neglected to disclose that its disproportionate holdings of long-term treasuries overexposed it to federal interest rate hikes.[90] If investors find something suspicious, they will bring a suit against the corporation. Those suits are a signal to the SEC that scrutiny may be warranted.

Resource- and information-strapped privacy enforcers could set up a structurally similar vetting process to help identify privacy violations. The first step would be to set up a system for mandating, filing, and publishing corporate privacy disclosures. Even the SEC has begun requiring businesses to file disclosures on some technological aspects of business operations. For example, the agency is in the process of implementing rules "to enhance and standardize disclosures regarding cybersecurity risk management, strategy, governance, and incident reporting."[91] Unfortunately, privacy enforcers probably cannot bootstrap themselves to the SEC's filing system. SEC disclosures are for information that materially relates to corporate financial performance. While cybersecurity risk counts — responding to and cleaning up after hacks is expensive — fidelity to privacy disclosures is not there yet.

Instead, privacy enforcers should establish their own filing system. As noted above, California authorities already have a nascent system in place: the CCPA requires data brokers to register basic information that the state then makes publicly available. The FTC could leverage its investigative powers to even broader effect.[92] The FTC Act authorizes the FTC to require corporations to "file with the Commission in such form as the Commission may prescribe annual . . . reports or answers

89. *Filings and Forms*, SEC. & EXCH. COMM'N (Jan. 9, 2017), https://www.sec.gov/edgar [https://perma.cc/C8UZ-8BTK].

90. Michelle Chapman & Associated Press, *Shareholders File Class Action Suit Against Silicon Valley Bank, Former CEO and CFO for Not Disclosing Rate Risk from June 2021 Through Its Collapse*, FORTUNE (Mar. 14, 2023), https://fortune.com/2023/03/14/silicon-valley-bank-svb-class-action-shareholder-lawsuit-risks-undisclosed-understated [https://perma.cc/D89R-W4J2].

91. SEC. & EXCH. COMM'N, FACT SHEET: PUBLIC COMPANY CYBERSECURITY; PROPOSED RULES, https://www.sec.gov/files/33-11038-fact-sheet.pdf [https://perma.cc/A8QP-PF2G].

92. We are grateful to Olivier Sylvain for pointing out this possibility.

in writing to specific questions."[93] In the past, the FTC has used this power to require broad disclosures from entire industries, like Internet Service Providers, social media companies, and streaming platforms.[94] A similar demand might be issued to corporations that handle larger quantities of consumer data, with specific questions that take shape as the forms we propose in the next section.

Once privacy enforcers receive and publish corporate privacy disclosures, their next step would be to find a way to engage external auditors. It must work differently than the SEC's system since, unlike investors who dissect SEC filings, there are currently few financial incentives for validating privacy disclosures.[95] Fortunately, privacy enforcers have partners ready and waiting.[96] Philanthropically motivated privacy activists and scholars already perform closed book audits, albeit slowly and at great expense. Privacy enforcers could streamline this work and significantly lower audit costs through standardized privacy disclosures. By creating mandatory disclosure templates for firms to use, privacy enforcers could pave the way for activists and scholars to develop automated systems that conduct closed book audits. The FTC might even adopt similar techniques itself.

B. Privacy Disclosure Forms

Privacy enforcers would need to construct their disclosure forms with a careful eye to their intended purpose: facilitating public accountability and closed book audits. The forms would need to convey relevant information in a manner that minimizes room for ambiguity and

93. 15 U.S.C. § 46(b).

94. FED. TRADE COMM'N, A LOOK AT WHAT ISPS KNOW ABOUT YOU: EXAMINING THE PRIVACY PRACTICES OF SIX MAJOR INTERNET SERVICE PROVIDERS ii (Oct. 21, 2021), https://www.ftc.gov/system/files/documents/reports/look-what-isps-know-about-you-examining-privacy-practices-six-major-internet-service-providers/p195402_isp_6b_staff_report.pdf [https://perma.cc/3ZF7-DJHZ]; *FTC Issues Orders to Social Media and Video Streaming Platforms Regarding Efforts to Address Surge in Advertising for Fraudulent Products and Scams*, FED. TRADE COMM'N, (Mar. 16, 2023), https://www.ftc.gov/news-events/news/press-releases/2023/03/ftc-issues-orders-social-media-video-streaming-platforms-regarding-efforts-address-surge-advertising [https://perma.cc/QQM4-FTMJ].

95. Financial incentives would be very different if private parties could bring class action suits for privacy violations. *See* Ignacio Cofone, *Certifying Privacy Class Actions*, 37 HARV. J.L. & TECH. 1147 (2024). Similarly, different incentives would be generated if the FTC offered a whistleblower bounty for discovering violations. *See* Ethan Hayward, *The Federal Government as Cookie Inspector: The Consumer Privacy Protection Act of 2000*, 11 DEPAUL-LCA J. ART & ENT. L. 227, 233 (2001) ("Apparently, whistleblowers are necessary in order for privacy violations to come to the attention of industry regulators."); *see also* Ying Hu, *Individuals as Gatekeepers Against Data Misuse*, 28 MICH. TECH. L. REV. 115, 132 (2021) ("[We should] incentivize whistleblowing by individuals who have first-hand knowledge of false declarations by data recipients.").

96. *See* David Choffnes, Woodrow Hartzog, Scott Jordan, Athina Markopoulou & Zubair Shafiq, *A Scientific Approach to Tech Accountability*, 37 HARV. J.L. & TECH. 1199 (2024).

in a format that is amenable to rapid or even automated processing. Once again, SEC disclosures offer a helpful starting point.

There are dozens of SEC forms, covering a wide range of formats and types of information. Not all of them are suitable to the present context. Consider, for example, Form 10-K, which publicly traded corporations must complete annually to provide a broad report on their financial performance.[97] The form has the advantage of providing instructions about what information to disclose, in what order, and with what section headings. This gives an overall structure to 10-K disclosures that is presently lacking from CCPA privacy disclosures. However, some parts of Form 10-K require corporations to report information in a manner that is inherently open-textured and narrative. Item 7, for example, is where the corporation provides "Management's Discussion and Analysis of Financial Condition and Results of Operations."[98] Item 7 reports are long,[99] highly variable,[100] and full of multiply interpretable language.[101] This is what one would expect. Publicly traded companies face idiosyncratic and evolving commercial environments that depend, among other factors, on the nuances of their line of business, geography, supply chain, credit supply, and shifting customer preferences. Forecasting financial performance is equal parts science and art.[102]

Fortunately, the broad contours of firms' data-handling practices are not quite so amorphous or idiosyncratic. The information that the CCPA requires in privacy disclosures is rather formulaic. The "Methods" standard, for example, simply requires covered businesses to specify two methods for consumers to submit a request to exercise a privacy right. There are a limited number of rights and an equally limited number of recognized modes of communication. Part III shows that the hodgepodge of disclosures firms presently make under the "Methods" standard are needlessly, obstructively varied.

---

97. SEC. & EXCH. COMM'N, FORM 10-K (2023), https://www.sec.gov/files/form10-k.pdf [https://perma.cc/FD6F-RM8L].

98. *See, e.g.*, Meta Platforms, Inc., Annual Report (Form 10-K), 54–76 (Feb. 2, 2023), https://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/e574646c-c642-42d9-9229-3892b13aabfb.pdf [https://perma.cc/3WJ9-THJ8].

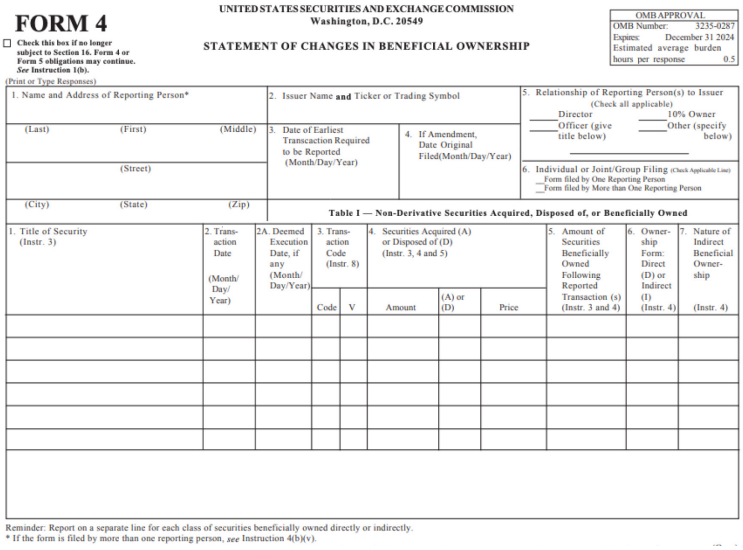
99. Facebook's last report was twenty-four pages. *Id.*

100. Jeffrey N. Gordon, *The Rise of Independent Directors in the United States, 1950–2005: Of Shareholder Value and Stock Market Prices*, 59 STAN. L. REV. 1465, 1553 (2007) (noting that item 7 disclosures "provide a narrative account of the financial results").

101. Amy Borrus, *The SEC: Cracking Down on Spin*, BLOOMBERG BUSINESSWEEK (Sept. 26, 2005), https://www.bloomberg.com/news/articles/2005-09-25/the-sec-cracking-down-on-spin [https://perma.cc/3YXY-ABQU].

102. Henry T. C. Hu, *Too Complex to Depict? Innovation, "Pure Information," and the SEC Disclosure Paradigm*, 90 TEX. L. REV. 1601, 1602 (2012) ("[Risk-return analysis of asset-backed securities] can be so complex that even 'objective reality' is subject to multiple meanings. Given such rudimentary tools and such complex realities, the depictions may offer little more than shadowy, gross outlines of the objective reality, however that reality might be conceived.").

The SEC does have forms for disclosing more routine types of information. Consider, for example, Form 4, which corporate executives use to report material changes in their personal investments in their own company's stock. One goal of the form is to help the SEC keep tabs on potential violations of insider trading laws. All the information about executive trades fits into cells of a simple chart. Indeed, its format is so predictable that third parties already have automated systems that scrape information from Form 4 and trade stock based on it.[103]



FORM 4

UNITED STATES SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

STATEMENT OF CHANGES IN BENEFICIAL OWNERSHIP

OMB APPROVAL
OMB Number: 3235-0287
Expires: December 31 2024
Estimated average burden hours per response 0.5

☐ Check this box if no longer subject to Section 16. Form 4 or Form 5 obligations may continue. *See* Instruction 1(b).

(Print or Type Responses)

1. Name and Address of Reporting Person*
(Last) (First) (Middle)
(Street)
(City) (State) (Zip)

2. Issuer Name **and** Ticker or Trading Symbol

3. Date of Earliest Transaction Required to be Reported (Month/Day/Year)

4. If Amendment, Date Original Filed(Month/Day/Year)

5. Relationship of Reporting Person(s) to Issuer (Check all applicable)
___Director ___10% Owner
___Officer (give title below) ___Other (specify below)

6. Individual or Joint/Group Filing (Check Applicable Line)
__Form filed by One Reporting Person
__Form filed by More than One Reporting Person

Table I — Non-Derivative Securities Acquired, Disposed of, or Beneficially Owned

1. Title of Security (Instr. 3)	2. Transaction Date (Month/Day/Year)	2A. Deemed Execution Date, if any (Month/Day/Year)	3. Transaction Code (Instr. 8)		4. Securities Acquired (A) or Disposed of (D) (Instr. 3, 4 and 5)			5. Amount of Securities Beneficially Owned Following Reported Transaction (s) (Instr. 3 and 4)	6. Ownership Form: Direct (D) or Indirect (I) (Instr. 4)	7. Nature of Indirect Beneficial Ownership (Instr. 4)
			Code	V	Amount	(A) or (D)	Price			

Reminder: Report on a separate line for each class of securities beneficially owned directly or indirectly.
* If the form is filed by more than one reporting person, *see* Instruction 4(b)(v).

Figure 4: SEC Form 4.[104]

A form tailored to the CCPA's Methods disclosure requirement could be equally straightforward. It might start by requiring basic identifying information for the company and, for each CCPA right, information about the method the firm makes available for customers to submit exercise requests.

103. Charles R. Korsmo, *The Audience for Corporate Disclosure*, 102 IOWA L. REV. 1581, 1592–93 (2017).

104. SEC. & EXCH. COMM'N, FORM 4 (2024), https://www.sec.gov/files/form4.pdf [https://perma.cc/6PSJ-QDPS]. Thanks to Robert Miller for suggesting Form 4 as an example.

CCPA Methods Disclosure Form

Business Name	*State of Incorporation*	*IRS Employee Identification Number*
Address of Principal Executive Officers		*Contact Phone Number*
CCPA Right	**Method (check all that apply)**	**Detail**
Right to Delete	*Phone*	*Phone Number*
	Email	*Email Address*
	Webform	*URL*
	Fax	*Fax Number*
Right to Correct	*Phone*	*Phone Number*
	Email	*Email Address*
	Webform	*URL*
	Fax	*Fax Number*

Figure 5: Proposed Sample Form for CCPA Disclosure of Methods for Consumers to Submit Requests to Exercise Privacy Rights

As Figure 5 illustrates, privacy disclosure forms need not be complex or nuanced, because the information they must convey is not complex or nuanced. The CCPA Methods disclosure requirement is just one example, but there is no reason to expect that the other disclosure requirements (e.g., what categories of personal information a firm collects) would be materially more challenging to reduce to a form. Perhaps the most difficult disclosures to formalize would be those that describe the categories of personal information that a firm collects and discloses. Yet even here, a standard template is not difficult to envision. It might include checkboxes next to a list of standard categories of information that firms often collect (e.g., location, IP address, email address) as well as an open text box where firms can also describe "Other" categories of information they collect.

While there are some similarities between the formulaic privacy disclosures we propose and a voluntary P3P[105] initiative in Europe, our model avoids the latter's disqualifying shortcomings. P3P was a short-lived protocol in the early 2000s that allowed websites to communicate their data management practices to a consumer web browser. The P3P standard failed for two reasons: it increased the burden on consumers and its violation carried no consequences. Unlike P3P, our proposal does not require consumers to install software or make sense of the formulaic disclosures. After all, consumers are not the target of our disclosure forms — regulators and auditors are. This decision also addresses the second major shortcoming of P3P. Because P3P disclosures were made to consumers, inaccuracies carried few legal

105. *See generally* Lorrie Faith Cranor, *P3P: Making Privacy Policies More Useful*, IEEE SEC. & PRIV., Nov.–Dec. 2003, at 50–52.

consequences. By contrast, misrepresentations in a disclosure to a government authority can entail substantial civil or criminal liability. Taken together, we expect that our proposal will facilitate more effective closed book audits by forcing accurate and standardized privacy disclosures while avoiding any increased burden on consumers.

V. CONCLUSION

Recent developments in privacy regulation have the potential to constrain big tech's data practices. But regulations are only as effective as regulators' enforcement of them, and enforcement is only possible when violations can be detected. Abusive data practices usually occur unobtrusively, behind the scenes. Fortunately, technical advancements in measurement and privacy research hold promise for uncovering covert mishandling of personal data. Privacy researchers have developed techniques for conducting closed book compliance audits that proceed without any cooperation from corporate targets. By streamlining or automating such audits, privacy regulators could open an enforcement window into hidden channels of the data ecosystem.

In this Essay, we investigated the privacy policies of California-registered data brokers to understand whether the CCPA's disclosure mandates are amenable to automated closed book audits. Unfortunately, our results show that there are still many challenges to overcome. Privacy policies today are hard to find, unstructured, vague, and unnecessarily verbose, even where they respond to simple disclosure requirements. We argued privacy enforcers could improve privacy disclosures by borrowing mandatory corporate disclosure practices from other areas of law. Formulaic disclosures may finally empower consumers and simultaneously unleash the power of automated closed book auditing systems.