Stanford University
Human-Centered
Artificial Intelligence

RegLab

Policy Brief
HAI Policy & Society
Stanford RegLab
August 2026

Regulating Data Brokers in the Age of AI: A California Case Study

Anna-Maria Gueorguieva, Jennifer King, Apoorva Panidapu, and Daniel E. Ho

Technology, social media, and AI mediate our daily lives, raising urgent concerns about how businesses collect, share, and sell our data. This collection process is largely invisible, occurring in our background use of apps, websites, and devices, without meaningful notice. The scale of this data collection is staggering: Billions of data points are generated and distributed each day. Data brokers — third-party companies that collect and exchange consumer data with whom they do not have first-party relationships — are key players in this opaque data-sharing ecosystem. The data that brokers make available for a price can include personal addresses, phone numbers, credit history, as well as predictive profiles assessing an individual's purchasing habits, insurance risk, and much more. All of this data can be used in ways that have harmful downstream effects for consumers, yet the broker ecosystem remains largely opaque.

Key Takeaways

Data brokers are third parties that buy and sell consumers' data without their specific knowledge or consent. In so doing, they expose consumers and the public to potential risks such as data breaches, political violence, and national security threats. Despite these risks, few states regulate this opaque ecosystem.

Widely lauded as the most comprehensive state consumer privacy statute, the California Consumer Privacy Act (CCPA) and the Delete Act require data brokers to allow consumers to exercise their privacy rights (e.g., by deleting their data or opting out of data collection) and mandate that brokers disclose the number of requests they receive each year.

We assess data broker compliance with the CCPA and the Delete Act, finding that many brokers obstruct consumers from properly exercising their requests and ignore disclosure requirements, leaving consumers to navigate a complex system with no direct remedy for harm.

Data brokers and generative AI developers operate in the same data ecosystem, with brokers selling consumer data to AI companies, making it imperative that data privacy protections extend to generative AI development.

In our paper "Privacy Without Remedy: An Assessment of Data Broker Compliance with California Privacy Law," we assess data broker compliance with the California Consumer Privacy Act and the Delete Act. California is widely seen to be at the frontier of comprehensive consumer privacy law, and these two acts are the first in the nation to require data brokers to: register with the state, allow consumers to exercise their data privacy rights, and publicly post the annual number of requests they receive from California consumers. We find that a majority of brokers ignore mandated disclosure requirements and add friction to rights request processes, making it challenging for consumers to exercise their rights.

A majority of brokers ignore mandated disclosure requirements and add friction to rights request processes.

These findings also matter in the context of generative AI development as novel datasets become ever more valuable for training AI systems. As the 2026 California data broker registry demonstrates, 32 brokers currently sell data to generative AI developers. To ensure that consumers can actively exercise their data privacy rights and that regulators and researchers can adequately monitor and assess data laws, policymakers should consider implementing similar data broker registration laws across the nation. Such legislation best serves consumers and researchers if it includes automated privacy rights request processes, standardized reporting practices, and the ability for consumers to pursue a private right of action.

Introduction

Most U.S. consumers are generally aware that their data is collected by the businesses they interact with directly. Fewer are knowledgeable about the wide-reaching third-party data collection ecosystem that exists alongside the websites and apps they visit regularly. Data brokers are the third-party companies that collect and sell the data of consumers with whom they do not have a first-party relationship. Some have over 10,000 different data types on offer for purchase by other businesses, individuals, and governments. Along with concrete behavioral data collected about consumers, brokers also make and sell *inferences* about future behavior — such as whether a consumer is likely to buy a product, take up a hobby, or need a specific service.

Many use cases for this data can be beneficial or benign, such as public health research, fraud detection, and identity verification. However, broker data has also been used in instances of political violence and to identify military personnel. For example, in 2020, a disgruntled lawyer used publicly available data available online to target District Judge Esther Salas and fatally shoot her son outside their New Jersey home. Broker data is also used to shape financial decisions that can have allocative harms.

In 2024, General Motors was found to be selling their customers' driving data without explicit consent to data brokers, who in turn used the data to create risk scoring products for auto insurance companies. The sales were discovered after many GM auto owners experienced rate increases, were dropped from their insurance policies, or were unable to find new auto insurance.

Such threats have fueled calls to regulate data brokers for over a decade, yet only four states currently do so. Under the California Consumer Privacy Act (CCPA) of 2018, data brokers must allow consumers to exercise their privacy rights to have their data deleted, corrected, and not sold to others. In addition, brokers must disclose what personal information is being collected, sold, or shared, and to limit the use and disclosure of sensitive personal information when requested by consumers. California's 2023 Delete Act requires brokers to publicly post the number of requests they receive from consumers in the prior calendar year. The goal of this requirement is to increase accountability by crowdsourcing compliance through the direct disclosure of information to the public. However, to date no other researchers have studied brokers' compliance with either of these laws.

Several data brokers also sell data to companies building generative AI, information that is newly disclosed in the 2026 California broker registry. Little is known regarding how generative AI developers might use broker-purchased data, and there are no transparency requirements that mandate any detailed disclosure.

Research Outcomes

We found low compliance with both the CCPA and the Delete Act, obstructing consumers' ability to exercise their rights and creating difficulty for researchers and policymakers to audit and enforce these laws.

Our study analyzed the 522 data brokers registered with the California Privacy Protection Agency in 2025. To analyze brokers' compliance, we visited and manually reviewed the websites of all 522 registered data brokers, verifying that each broker had a posted privacy policy and whether they reported rights requests, both before and after the July 1, 2025, compliance deadline. We found that 45% did not report any information about requests received, and only 9.2% were fully compliant with reporting information for all rights request types (deletion, opt-out, know what is being collected/sold, and limiting sharing sensitive information). Among brokers that do report rights request metrics, fulfillment rates exceeded 80% for most request types, except for "right to know" requests, which have a 73% fulfillment rate.

Thus, while some data brokers are generally complying with consumer rights requests, a majority of data brokers are not compliant with mandatory disclosure and do not provide enough information to allow researchers and regulators to systematically assess their compliance. Additionally, collecting data across data brokers is laborious, requiring individual visits to each broker's online privacy policy (as we did), where rights requests must be included or linked. While this requirement aims to increase public awareness, the decision to allow brokers to self-report makes systematic monitoring costly, even with a known, identified population of registered data brokers.

We also examined consumer request processes of a subset of 250 data brokers and found direct violations of the CCPA and interface features that, while not in direct violation, create considerable friction for users. This step also required manual review and coding of the online interface for each broker's rights request processes. We found significant friction with the processes for consumers to exercise their privacy rights. Because the CCPA relies on an opt-out framework — as opposed to an opt-in — the burden of exercising privacy rights falls on consumers who must successfully navigate non-standardized and sometimes unnecessarily complex online rights request processes. The design and usability of these processes are thus crucial to the efficacy of the law in practice.

In terms of CCPA violations, 43% of sampled brokers did not support all six consumer rights requests. Additionally, 22% of brokers require a CAPTCHA submission in request processes, an extraneous feature to privacy rights requests; in one extreme case, a broker required consumers to fill out eight different CAPTCHAs, including math problems. These design choices matter because they affect whether consumers can exercise their privacy rights in practice. When request processes are fragmented across hundreds of brokers, features such as repeated forms, broken links, unnecessary or overly burdensome verification steps make it less likely that consumers will successfully exercise their privacy rights. In comparing the number of consumer requests to the amount of friction identified in brokers' request processing, we found that data brokers with less interface friction received more requests.

Altogether, these findings show that privacy rights are most meaningful and effective when the consumer processes for exercising them are accessible, consistent, and friction-free.

Privacy rights are most meaningful and effective when the consumer processes for exercising them are accessible, consistent, and friction-free.

Policy Discussion

Data brokers contribute to a wide-reaching third-party data collection ecosystem that has negative downstream impacts for consumers. Our work shows that even under California's relatively strong consumer privacy laws, most brokers fail to comply with transparency requirements, and their request processes introduce substantial friction for consumers, making it challenging for them to exercise their rights. We present five policy recommendations to increase the efficacy of these regulations. Our recommendations are also relevant to AI development, as brokered data is becoming part of the AI data supply chain.

1. Require broker registration. While the California data broker registry is self-selected, with limited monitoring and penalties for brokers that do not register, our study was only possible because this registry exists. Identifying data brokers beyond the largest ones is a challenge. Requiring data broker registration across

states makes an otherwise opaque ecosystem visible to consumers, researchers, and regulators. The current proposed federal privacy bill, SECURE Data Act (H.R. 8413), includes a similar requirement for data broker registration. This is a good idea, providing a starting point for understanding the web of data brokers, as is the California 2026 requirement to indicate if brokers sell to generative AI companies.

2. Adopt automated and scalable privacy rights request processes. Brokers with less friction in their rights request processes handle more consumer requests. We encourage regulators to mandate a set of standardized rights request templates to make the request process uniform. However, it is not only friction that makes it difficult to exercise requests. Consumers must make rights requests on a per-company basis, a tedious task that most consumers will not have the time or patience to complete across potentially hundreds of brokers. One solution is to provide automated and scalable mechanisms to exercise rights across a set of businesses, including data brokers. For example, California's DROP (Delete Request and Opt-out Platform) provides consumers with a one-stop shop for opting out of data collection from registered data brokers. Since its launch in January 2026, over 300,000 Californians have requested to delete their data and opt out of broker data collection. While much debate exists around whether data collection defaults should be opt-in vs. opt-out, automated approaches like DROP will allow consumers to more easily exercise their rights at scale, enabling them to reduce their exposure to data brokers and AI development.

3. Standardize reporting practices. Requiring public posting of request metrics enables greater knowledge sharing and oversight of data privacy practices by businesses. Our research showed that obtaining this information from online privacy policies is a laborious and unscalable task. To allow consumers, regulators, and researchers the full benefit of disclosure requirements, regulations should require that rights request metrics be presented in both standardized human and machine readable formats. Our analysis also found that the majority of brokers posted metrics that were at best incomplete and, in some cases, likely incorrect when compared to their metrics submitted to California's registry. Standardized reporting practices can help address some of these gaps and make it easier to assess whether data privacy laws are functioning as intended.

4. Include a private right of action. Given constraints on enforcement resources, there are strong reasons to update the CCPA to include a private right of action to at least secure compliance with transparency provisions and/or allow aggregation of claims against data brokers for violations of the law. Public enforcement remains necessary, but agency resources are limited and have clearly led to suboptimal transparency. Outside of the context of California, no other states that currently regulate data brokers include a private right of action. Considering the uneven compliance we document in California, it is likely that brokers in other states are also not complying with their respective laws. A lack of private enforcement combined with constraints on public resources means that broker compliance will likely remain weak.

5. Extend protections to AI contexts. California's 2026 registry now requires brokers to disclose whether they sell data to generative AI companies, reflecting the growing concern about the potential data privacy impacts of these tools. AI development complicates data protections, enhancing downstream risks of breaching data privacy and contributing to the already

opaque ecosystem of consumer data collection and processing. As data grows increasingly valuable for training massive AI systems, consumer data privacy law must also account for the interactions between AI developers and third-party intermediaries that can collect, infer, sell, and circulate personal data. Too little is known about the AI supply chain, and broker transparency is a critical piece.

Historically, data brokers have gone understudied in the context of consumer privacy.

Conclusion

Historically, data brokers have gone understudied in the context of consumer privacy. One of the strongest aspects of California's privacy law framework is requiring data broker registration, making an otherwise opaque ecosystem visible to consumers, researchers, and regulators. However, despite the increasing participation in California's recent DROP program — indicating that such request processes are desired and beneficial to consumers — our research reveals a persistent lack of transparency and compliance from data brokers. As other states increasingly adopt privacy laws that follow California's example, reporting requirements must be sufficiently standardized to support regulatory oversight, and privacy rights request processes must be simple and straightforward enough so consumers actually use them. Importantly, privacy laws must provide meaningful remedies when those rights are obstructed to accomplish what they were written for: to enable consumers to reclaim agency over their data.

Reference: The original article is accessible at Anna-Maria Gueorguieva et al., "**Privacy Without Remedy: An Assessment of Data Broker Compliance with California Privacy Law**," ACM Conference on Fairness, Accountability, and Transparency (FAccT), June 2026, https://dl.acm.org/doi/10.1145/3805689.3812413.

Stanford University's Institute for Human-Centered Artificial Intelligence (HAI) applies rigorous analysis and research to pressing policy questions on artificial intelligence. A pillar of HAI is to inform policymakers, industry leaders, and civil society by disseminating scholarship to a wide audience. HAI is a nonpartisan research institute, representing a range of voices.

Stanford University's Regulation, Evaluation, and Governance Lab (RegLab) partners with government agencies to design and evaluate programs, policies, and technologies that modernize government.

The views expressed in this policy brief reflect the views of the authors. For further information, please contact **HAI-Policy@stanford.edu**.

Anna-Maria Gueorguieva is a PhD student at the University of Washington Information School and was a Summer Institute Fellow at the Regulation, Evaluation, and Governance Lab (RegLab) at Stanford University.

Jennifer King is the Privacy and Data Policy Fellow at the Stanford University Institute for Human-Centered Artificial Intelligence (HAI).

Apoorva Panidapu is an undergraduate researcher at the RegLab at Stanford University.

Daniel E. Ho is the William Benjamin Scott and Luna M. Scott Professor of Law, Professor of Political Science and Computer Science (by courtesy), Associate Director and Senior Fellow at the Stanford HAI, Senior Fellow at the Institute for Economic Policy Research (SIEPR), and director at the RegLab.

Acknowledgments

We are grateful to Caroline Yee and Jason Shin for their research assistance with this project, and to the summer 2025 Stanford RegLab cohort for their suggestions.

Stanford University
Human-Centered
Artificial Intelligence

Stanford HAI: 353 Jane Stanford Way, Stanford CA 94305-5008

T 650.725.4537 **F** 650.123.4567 **E** HAI-Policy@stanford.edu **hai.stanford.edu**