Perspectives on Amending the Publicly Available Information Exemption of the California Privacy Rights Act

By Samuel Leitch

The California Privacy Rights Act (CPRA) and its predecessor, the California Consumer Privacy Act of 2018 (CCPA), are some of the most comprehensive data privacy laws in the United States. Both laws broadly establish the rights that Californians possess regarding their personal data and what companies can do with it. These rights include, among others, the right to correct inaccurate information, the right to have personal information deleted, and the right to opt out of having personal information sold to or shared with third parties.[1]

When the California State Legislature passed the CCPA in 2018,[2] some claimed that the bill included "one of the broadest definitions of personal information under U.S. law."[3] It exempted two main categories of data: information "lawfully made available from federal, state, or local government records" and "deidentified or aggregate consumer information."[4] In amending the CCPA, however, the CPRA narrowed this definition through a major exemption: publicly available information, even when it reveals sensitive details about an individual, is exempt under the CPRA.[5]

Web scraping has facilitated the automated extraction of this publicly available information from the Internet. While bots can only access the same websites as any human with an Internet connection, they do so at a rate that would take a human several lifetimes, collecting phone numbers, profile pictures, blog posts, and more along the way.

---

1 California Consumer Privacy Act, Privacyrights.org (2020), privacyrights.org/resources-tools/law-overviews/california-consumer-privacy-act.
2 Id.
3 Erin Illman & Paul Temple, California Consumer Privacy Act: What Companies Need to Know, 75 The Business Lawyer 1637 (2019), www.jstor.org/stable/27171063.
4 Cal. Civ. Code § 1798.145(c)-(f), 1798.185(a)(3).
5 Cal. Civ. Code § 1798.140(v)(2)(B).

The expanding market for artificial intelligence "has increased the motivation to scrape, as AI demands vast amounts of training data."[6] One example is when a startup named Clearview AI acquired over three billion images across the web, often in violation of terms of service, to train its proprietary facial recognition algorithm.[7] At least 600 law enforcement agencies currently lease Clearview AI's algorithm for a fee.[8] Customs and Border Protection finalized a deal with the company earlier this year: for an annual cost of $225,000, the agency gains access to a database that now contains over 60 billion publicly available images.[9] Since Clearview AI only used publicly available photographs to train its algorithm, its database is exempt from the CPRA. However, while these images may have been publicly available, the users who posted them never consented to being included in a facial recognition database—especially not the individuals "whose photos were publicly available without their knowledge or consent."[10]

In this era of online platforms, users often take for granted, as a premise barely worth considering, that the Internet is never really private. Conventional wisdom warns against posting anything that an employer or admissions council might find objectionable. Nevertheless, even the most cautious of California residents will make their information public. A user might post her phone number and home address to LinkedIn in search of a job. Her current employer might also, without her knowledge, upload her headshot to the company's staff page. The CCPA and CPRA allow her to delete this information if someone collects it, but it is difficult to imagine that she would exercise this right: scrapers do not need to notify users before collecting their data. Consequently, as machine learning algorithms both facilitate and create demand for the automated mass extraction of publicly available personal information, the CPRA may prove ineffective in protecting Californians' privacy rights.

---

6 Id.

7 Daniel Solove & Woodrow Hartzog, The Great Scrape: The Clash Between Scraping and Privacy, 113 California Law Review (2024), papers.ssrn.com/sol3/papers.cfm?abstract_id=4884485.

8 Kashmir Hill, Unmasking a Company That Wants to Unmask Us All, The New York Times, Jan. 20, 2020, www.nytimes.com/2020/01/20/reader-center/insider-clearview-ai.html.

9 Dell Cameron, CBP Signs Clearview AI Deal to Use Face Recognition for "Tactical Targeting," WIRED (2026), www.wired.com/story/cbp-signs-clearview-ai-deal-to-use-face-recognition-for-tactical-targeting/.

10 Katelyn Ringrose & Divya Ramjee, Watch Where You Walk: Law Enforcement Surveillance and Protester Privacy, 11 California Law Review Online 360 (2020).

Even if Americans want to protect their privacy, the majority have no idea how: almost 70 percent of Americans claim to know "little to nothing" about what companies actually do with their personal data, and 81 percent are concerned about this usage. It comes as no surprise, then, that there is bipartisan support nationwide to strengthen regulations on how companies handle personal data: 78 percent of Democrats and 68 percent of Republicans express this opinion, and only 7 percent say there should be less regulation.[11] This support extends globally. In 2018, the European Union passed its General Data Privacy Regulation (GDPR), requiring companies to obtain consent from consumers before using their data, to explain the scope of this usage to consumers, and to alert consumers in the event of a data breach.[12] Since then, several countries across the globe have adopted GDPR-style privacy regulations.[13]

However, any change to California's privacy legislation would have major consequences. For one, any amendment that heavily impedes web scraping could make the Internet unusable. Web scraping was "used to build what we know as the World Wide Web," and to this day it "enables web searching, archiving, generative AI, and scientific research."[14] Without web scraping, the Internet would be much harder to navigate, if not impossible. Furthermore, California's privacy legislation is not only relevant to California residents. Its effects "are on a national and international scale as it affects any company who does business with consumers in California."[15] Consequently, amending the publicly available information exemption would affect advertisers, journalists, and researchers, both domestically and abroad, who rely on publicly available information to varying degrees.

---

11 Michelle Faverio, Key Findings about Americans and Data Privacy, Pew Research Center (2023), https://www.pewresearch.org/short-reads/2023/10/18/key-findings-about-americans-and-data-privacy/.
12 Ronan Murphy, Mapping the Brussels Effect: The GDPR Goes Global, CEPA (2025), https://cepa.org/comprehensive-reports/mapping-the-brussels-effect-the-gdpr-goes-global/.
13 Id.
14 Solove, supra note 7, at 1523.
15 Jacklin Lee, CCPA/CPRA: Consumers Bear the Burden as Companies Bear the Crown, 47 UC Law SF International Law Review 130 (2024), repository.uclawsf.edu/cgi/viewcontent.cgi?params=/context/hastings_international_comparative_law_review/article/1891/&path_info=4_CCPACPRA__US_Europe___privacy_law_.pdf.

Journalists in particular might currently benefit from the CPRA's exemption of personal information that appears in "widely distributed media."[16] Without this exemption, subjects of newspaper articles could possibly use specious privacy-related claims to delete any quotes or facts that portray them in a bad light. Even if such claims were not to prevail in court, they would possibly exert considerable strain on newsrooms throughout America.

Considering the complexity of this issue, this brief aims to collect and present arguments both for and against amending the publicly available information exemption with the goal of guiding further discussion on statewide regulation.

What is third-party data?

Big data has revolutionized the advertising industry. Back when advertisements were a novel idea, advertisers had to rely on direct polls and surveys to judge the effectiveness of a given campaign—methods that required the express consent of the data subject.[17] Through online applications and web browsers, however, advertisers can poll users passively, automatically, and at a level of specificity previously considered impossible.

The resulting trove of user data falls into four possible categories.[18] Zero-party data most closely mirrors traditional polling. A user actively shares zero-party data with a website: for example, through a feedback survey.[19] First-party data also relates to a user's direct interaction with a website, but this data is collected, not given. Publishers and advertisers compile analytics like subscription info, transaction history, views, clicks, time spent on a particular page, and attitudinal data from likes, shares, and comments.[20]

---

16 Cal. Civ. Code § 1798.140(v)(2)(B).
17 Allan E. Holder, What We Don't Know They Know: What to Do About Inferences in European and California Data Protection Law, 35 Berkeley Technology Law Journal 1331 (2020).
18 Meaghan Donahue, What Do State Privacy Laws Mean for the Ad Tech Industry?, New America (2026), www.newamerica.org/insights/what-do-state-privacy-laws-mean-for-the-ad-tech-industry/ (last visited Apr 27, 2026).
19 Id.
20 Id.

When an advertiser or publisher shares this data with a business partner, it becomes second-party data.[21] Companies often provide application programming interfaces (APIs) that grant other businesses specific ways to view, query, and download these datasets. When a user loads a webpage, supply-side platforms may also broadcast this information across the Internet in an automated bidding war for ad space on the user's device.[22] These auctions, called ad exchanges or real-time bidding networks, are often over in mere milliseconds.[23]

The last category, third-party data, involves everything else: data collected by entities with no direct relationship to the user. The largest dealers are known as data brokers, or companies that "specialize in buying, aggregating, and selling the personal data of individuals with whom they have no direct relationship."[24] In 2021, the global data broker market was valued at $319 billion, and this number may reach $545 billion in 2028.[25]

The data brokerage ecosystem collects and sells information on nearly every American, information that includes "age, race, ethnicity, sex, gender, sexual orientation, religion… political preferences and beliefs… geolocations, health conditions… and lifestyle characteristics."[26] It is worth mentioning that this is the exact category of data that the CPRA considers "sensitive personal information." Some data brokers aggregate data profiles of hundreds of millions of people in the United States, and these profiles can range from "a few data points…to hundreds or thousands of data points about a single person."[27]

---

21 Id.

22 Jack Marshall, What is an Ad Exchange?, Digiday (2014), digiday.com/media/what-is-an-ad-exchange/.

23 Id.

24 Andy Z Wang, Network Harms, 91 The University of Chicago Law Review 2093 (2024).

25 Id.

26 Justin Sherman et al., Response from Duke University's Data Brokerage Research Project Consumer Financial Protection Bureau (CFPB) Request for Information Regarding Data Brokers and Other Business Practices Involving the Collection and Sale of Consumer Information, (2023), techpolicy.sanford.duke.edu/wp-content/uploads/2023/07/CFPB-2023-0020-3962_attachment_1.pdf (last visited Apr 27, 2026).

27 Id.

Data brokers collect this data directly, indirectly, and through inferences. Direct collection could involve, for example, "contracts with app developers to include the broker's data siphoning software directly in their apps."[28] In mobile applications, this may involve software development kits (SDKs) with tracking software. In the browser, this may involve third-party or first-party cookies.

Indirect collection may involve, on the other hand, scraping public records and online platforms, collecting user profiles from real-time bidding networks, and purchasing data from first-party collectors or other brokers.[29]

Lastly, data brokers use inferences, or predictions based on existing data. In a rather infamous example, Target's promotional algorithms correctly determined that a teenage girl was pregnant based on her purchases.[30] She was one of the thousands of entries labeled "most likely pregnant" in Target's national database.[31]

Inferences often blur the line between sensitive and non-sensitive information. As technology evolves, "sensitive information may also be able to be discerned from categories of information not traditionally thought of as sensitive," an example being the "emerging advancements in voice analysis that promise to discern an individual's COVID-19 status through the sound of their cough."[32] Earlier this year, Apple acquired a startup that uses "'facial skin micromovements' to detect words mouthed or spoken, identify a person and assess their emotions."[33] These inference-making technologies are only the tip of the iceberg, and California residents will likely see many more such advancements in the years to come.

---

28 Wang, supra note 24, at 2100.
29 Wang, supra note 24.
30 Charles Duhigg, How Companies Learn Your Secrets, The New York Times, Feb. 16, 2012, https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.
31 Id.
32 Katelyn Ringrose, New categories, new rights: The CPRA's opt-out provision for sensitive data, IAPP.org (2023), iapp.org/news/a/new-categories-new-rights-the-cpras-opt-out-provision-for-sensitive-data (last visited May 11, 2026).
33 Stephen Nellis, Apple acquires Israeli audio AI startup Q.ai, Reuters, Jan. 29, 2026, https://www.reuters.com/business/apple-acquires-audio-ai-startup-qai-2026-01-29/.

Should third-party businesses have free rein over personal information?

Third-party data sharing is a topic of vast complexity. Consumers often benefit from third parties owning their data: the practice "enables banks to more effectively detect fraudulent charges," provides information to cities "on how commuters travel via car, bike, bus, or rail," and "can be used to tailor advertising."[34] As mentioned in previous sections, web scraping, a major source of third-party data, is also vital for researchers, journalists, advertisers, and anyone who uses a search engine. Consequently, any regulation involving publicly available information—and, by extension, the data available to web scrapers and data brokers—will undoubtedly affect free speech and the usability of the modern Internet.

On the other hand, data brokers may endanger civil liberties, marginalized communities, and national security. It is currently possible for private individuals to purchase sensitive information from data brokers for surveillance purposes. From 2018 to 2021, a conservative nonprofit spent millions of dollars on app data from apps like Grindr to identify gay priests and report them to church leadership—resulting in the resignation of a Catholic bishop.[35] Acxiom, a broker with data on billions of people worldwide, claims it can determine "if someone has visited a location multiple times in the past 30 days, like a church, their therapist's office, or their ex's house."[36] By purchasing geolocation data from brokers like Acxiom through people-search websites, abusers can more easily stalk, harass, and assault intimate partners; women and LGBTQ+ individuals are most often the targets of this violence.[37] Meanwhile, notable broker LexisNexis, which advertises over 283 million profiles of American consumers, advertises "a capability to identify active military personnel," putting them at risk of foreign surveillance.[38]

---

34 Wang, supra note 24.
35 Michelle Boorstein & Heather Kelly, Catholic group spent millions on app data that tracked gay priests, Washington Post, Mar. 9, 2023, https://www.washingtonpost.com/dc-md-va/2023/03/09/catholics-gay-priests-grindr-data-bishops/.
36 Justin Sherman, Data Brokers Know Where You Are—and Want to Sell That Intel, WIRED (2021), http://www.wired.com/story/opinion-data-brokers-know-where-you-are-and-want-to-sell-that-intel/ (last visited May 11, 2026).
37 Sherman et al., supra note 26.
38 Id.

Government agencies often purchase information from data brokers to conduct warrantless investigations in violation of the Fourth Amendment. Through utility records purchased from data brokers like Thompson Reuters, Immigration Customs and Enforcement (ICE) can locate 3 in 4 adults, and it spent roughly $2.8 billion between 2008 and 2021 to expand its surveillance dragnet.[39] One data broker, SafeGraph, sells the location data of women who visit Planned Parenthood centers.[40] While California has passed a shield law to prohibit electronic communication providers from complying with investigations into reproductive activities, this law does not prevent data brokers from selling information to law enforcement agencies.[41] This information may be considered sensitive under the CCPA, but it is difficult to imagine that a user would exercise her right to delete personal information if she is unaware that the collection ever occurred.

For these reasons, the question of whether or not third-party businesses should retain access to personal information is exceedingly complex. Companies may argue that third-party data sharing between supply-side and demand-side platforms enables them to provide personalized deals and discounts to consumers. They may also argue that the regulation of personal information harms profit margins. Meta has previously claimed that the CCPA and other related laws have caused its advertising revenue to be "adversely affected by reduced marketer spending as a result of limitations on…ad targeting and measurement tools."[42] This brief will explore whether or not the right to delete publicly available personal information harms businesses in the following section.

On the other hand, some scholars advocate for a nuanced approach to third-party data sharing that mirrors the GDPR. Under Article 6 of the GDPR, a lawful ground is required to use or

---

39 American Dragnet | Data-Driven Deportation in the 21st Century, American Dragnet (2022), https://americandragnet.org.

40 Joseph Cox, Data Broker Is Selling Location Data of People Who Visit Abortion Clinics, VICE (2022), https://www.vice.com/en/article/location-data-abortion-clinics-safegraph-planned-parenthood/.

41 Jocelyn Frye, Digital Surveillance Supercharges Abortion Criminalization. Closing the Data Broker Loophole Is Urgent | National Partnership for Women & Families, National Partnership for Women & Families (2024), https://nationalpartnership.org/digital-surveillance-supercharges-abortion-criminalization-closing-data-broker-loophole-urgent/.

42 Morgan Carter, The Optimal Opt-In Option: Protecting Vulnerable Consumers in the Expanding Privacy Landscape, 124 Columbia Law Review 445 (2024), https://columbialawreview.org/content/the-optimal-opt-in-option-protecting-vulnerable-consumers-in-the-expanding-privacy-landscape/.

collect personal data.[43] When it comes to sensitive personal information, data controllers must meet one of the permissions under Article 9(2), "such as explicit opt-in consent, providing medical services, or for scientific research purposes, only as long as necessity and proportionality conditions are met."[44] The result is a multi-tiered system that diverges from the CPRA. The GDPR protects all personal data by default, adding more specific protections for sensitive personal data. While the CPRA and CCPA require Californians to opt out of third-party data sharing, Europeans opt out by default, and they must opt in before data sharing or processing can take place.

One paper argues that California should remove the exemption entirely, keeping "publicly available personal information within the scope of California's privacy protections, as it remains within the scope of the GDPR's protections."[45] The argument here is that California could ban most data scraping "while permitting innocuous collections of personal information" that are necessary to journalists and researchers, for example.[46] This would "be similar to permitting scraping where there is a lawful basis under the GDPR," which requires data controllers to meet one of the conditions provided under Article 9(2).[47] While many data brokers provide valuable services, it is not difficult to imagine that several data brokers would fail to meet this public interest requirement. Removing or altering the CPRA exemption would no doubt profoundly impact the data brokerage ecosystem.

Would the right to delete publicly available personal information harm businesses?

As discussed in the previous section, publicly available personal information in the form of consumer profiles enables the modern advertisement agency to cater to increasingly specific demographics. This practice of behavioral advertising, where advertisers purchase ad space

---

43 Ringrose, supra note 32.
44 Id.
45 Andrew Parks, Unfair Collection: Reclaiming Control of Publicly Available Personal Information from Data Scrapers, 120 Michigan Law Review 914 (2022), https://michiganlawreview.org/journal/unfair-collection-reclaiming-control-of-publicly-available-personal-information-from-data-scrapers/.
46 Id.
47 Id.

based on the individual characteristics of the user, differs from contextual advertising, where advertisers purchase ad space based on the content of the webpage, video, or application.[48] Some advertisers claim that behavioral advertising leads to a higher click-through rate and return on investment.[49]

If data brokers—and, by extension, the data management platforms and advertisers that use them—no longer had free rein over publicly available personal information, it is safe to say that this would limit the effectiveness of behavioral advertising. One study found that "the inability to behaviorally target opt-out users results in a loss of about $8.58 in ad spending per American opt-out consumer, which is borne by publishers and the exchange."[50] Some might argue that this would harm small websites by limiting the amount of capital available for purchasing ad space.

However, contextual advertising might have certain advantages over behavioral advertising, at least for publishers. Not wanting to lose subscribers by running afoul of the GDPR, the New York Times blocked all real-time bidding networks and behavioral targeting on its European webpages.[51] However, digital advertising revenue didn't fall; according to their reports, it "increased significantly."[52] As one of the largest journalistic enterprises globally, the New York Times might not represent all businesses, but this example casts some doubt on the superiority of behavioral advertising.

There is also evidence that even though behavioral ad targeting can cost twice as much as contextual advertising, it might not even be as effective.[53] One study found that targeting based on gender was only accurate 42 percent of the time—a percentage less than the natural

---

48 Donahue, supra note 18.
49 The Pros and Cons Behind Behavioral Marketing, Business Partner Magazine (2020), https://businesspartnermagazine.com/pros-cons-behind-behavioral-marketing/.
50 Garrett A. Johnson, Scott Shriver & Shaoyin Du, Consumer Privacy Choice in Online Advertising: Who Opts Out and at What Cost to Industry?, 39 SSRN Electronic Journal (2017).
51 Jessica Davies, After GDPR, The New York Times cut off ad exchanges in Europe—and kept growing ad revenue, Digiday (2019), https://digiday.com/media/gumgumtest-new-york-times-gdpr-cut-off-ad-exchanges-europe-ad-revenue/.
52 Id.
53 Augustine Fou, How Accurate Is Programmatic Ad Targeting?, Forbes (2020), https://www.forbes.com/sites/augustinefou/2020/09/10/how-accurate-is-programmatic-ad-targeting/.

population split.[54] This is to say that, at least in some cases, serving ads indiscriminately to all users could be more accurate than behavioral advertising. Many of the parameters that data brokers and data management platforms use for behavioral advertising are based on inferences, and these inferences, in the end, are guesses that can be incorrect.[55]

However, this analysis is not to downplay the impact of compliance costs. To analyze the economic effects of compliance, it may be appropriate to consider the GDPR as a case study. Article 14 of the GDPR requires data controllers to notify data subjects that they intend to process their data, even "when their personal data have been scraped off the public internet"; meanwhile, Article 5 ensures that even after scrapers have provided "such notice to all of the data subjects, the scraping must still meet certain criteria for it to be lawful."[56] If California amends the publicly available information exemption, then California businesses may incur expenses similar to those of their EU counterparts.

Some surveys estimate that compliance with GDPR can cost anywhere "from $1.7 million for small and midsize firms up to $70 million for larger ones, much of this being driven by the need to hire more employees and upgrade technology."[57] Another estimate suggested that "the GDPR was responsible for a 20 percent increase in the average cost of data."[58] This demonstrates that privacy regulations often come with tangible costs, and regulators must be cautious not to hinder American businesses, especially small businesses. However, it is also worth noting that the CCPA and CPRA only apply to certain businesses: those with over $25 million in annual revenue, those that primarily sell personal data, and those that handle large amounts of personal data.[59] Aside from small data brokers, any changes to these regulations would mostly affect large and mid-size transnational corporations, many of which have already invested in the necessary infrastructure to comply with the GDPR.

---

54 Id.
55 Id.
56 Parks, supra note 45, at 935.
57 Dylan Walsh, GDPR reduced firms' data and computation use, MIT Sloan (2024), https://mitsloan.mit.edu/ideas-made-to-matter/gdpr-reduced-firms-data-and-computation-use.
58 Id.
59 Cal. Civ. Code § 1798.140(d)(1).

If users of online platforms fail to use privacy controls, should their privacy rights be considered forfeited?

It could be argued that users of online platforms possess a great deal of agency over their privacy rights. In addition to the opt-out rights established by the CCPA and CPRA, online platforms allow users to set their profiles from public to private, thereby restricting the audience that can view their posts. However, if users fail to use these controls—either due to forgetfulness, a lack of understanding, or skepticism that these controls even matter—their information is public and therefore up for grabs. The following question, then, is whether or not this is a reasonable framework for consumers.

Whatever the case, when it comes to the complex web of services, platforms, and middleware that make up the World Wide Web, Internet users struggle with the decision to exercise their opt-out rights. More than half of Americans admit to always or almost always agreeing to privacy policies without reading them, and 61 percent are skeptical that their choice even matters.[60] A given user's decision to opt out or not is often about as deliberate as a coin toss. Even privacy researchers and industry experts can struggle with this decision. To completely understand the given risks, users "would need to speak at length with [the company's] chief privacy officer or data protection officer… review any algorithms that are making decisions about [them] as well as the data that the algorithms are being trained on, review their specific data security safeguards and how well they are being implemented… and on and on."[61] Even if users had the time to gather this information on one organization and weigh the risks, which is certainly not the case, it would require an incalculable amount of effort to do so for the thousands of websites that may ask for a given user's opt-out decision.

The question of opting out becomes even more difficult when companies offer consumers exclusive offers and loyalty programs in exchange for their data. For example, a person could be "asked to consent to allowing an online retailer to track her activity" in exchange for "a 10% discount at the store… the decision is not one that people spend months ruminating on; it is made

60 Faverio, supra note 11.

61 Daniel Solove, Murky Consent: An Approach to the Fictions of Consent in Privacy Law, 104 SSRN Electronic Journal 619-620 (2023).

quickly."[62] While the CPRA prohibits discrimination against users who opt out, it still permits a difference in price or service if that difference "is reasonably related to the value provided to the business by the consumer's data."[63] It is worth mentioning that this is one of the reasons that the Electronic Frontier Foundation did not support the CPRA, writing that it "would allow a business to withhold a discount from a consumer, unless the consumer lets the business harvest granular data about their shopping habits, and then profit on disclosure of that data to other businesses."[64] This arrangement, in which a business can influence a user to surrender certain privacy rights, is also known as a pay-for-privacy scheme.[65]

Businesses might argue, however, that these frameworks allow them to offer discounts to customers, lowering prices. In response to AB 446, a proposed bill that would have banned surveillance pricing, a coalition of businesses wrote that many companies might "stop offering discounts" and "choose to cancel even potentially compliant discounts because the cost of potential litigation and shakedown demand letters is too great."[66] It is likely that businesses would argue these points in response to similar changes in the publicly available information exemption.

Digital literacy is not uniform for all demographics, however, and marginalized communities that have historically lacked access to digital education might very well be facing disparate impact under the current opt-out system. For example, opt-out rates are very high for higher-income individuals and older populations, but that rate "falls with both the Asian- and African-American population shares."[67] The data suggests that while Asian-Americans are more likely to use ad blocker software, giving them some degree of protection against the trackers embedded in online

---

62 Id, at 620.
63 Cal. Civ. Code § 1798.125(a)(2).
64 Lee Tien, Why EFF Doesn't Support California Proposition 24, Electronic Frontier Foundation (2020), https://www.eff.org/deeplinks/2020/07/why-eff-doesnt-support-cal-prop-24 (last visited May 11, 2026).
65 Id.
66 Cal. S. Judiciary Comm., Analysis of A.B. 446, at 11 (2025).
67 Carter, supra note 42, at 448.

advertising, Black people are less likely to do so.[68] For this reason, Black people may disproportionately lack privacy protections.

Others may argue that the debate over opting out misunderstands the concept of user consent. A user might, for example, decide to make information public at some point and later withdraw that consent. The mere fact that "a user posts her home address on a publicly available website does not eliminate her interest in later preserving the privacy of that information."[69] This user "may have made the post public only temporarily," "accidentally posted it publicly," or posted it to a private profile where "one of those friends with access may have reposted or redistributed her information publicly."[70] There are also the many cases where an individual never knows that their personal information is publicly available: public records, data breaches, doxing, and revenge porn are all examples.[71]

For this reason, some argue that California should adopt an opt-in system similar to that of the GDPR. It is possible that the aforementioned "lack of knowledge by the lay consumer coupled with the opt-in default… make the CPRA less effective than it could be compared to the GDPR."[72] By requiring businesses to proactively acquire consent before collecting personal information, an opt-in framework "may allow for the less digitally literate of being one of the primary demographics left with their data traded by businesses without their understanding."[73] The CCPA itself recognizes the importance of an opt-in right for certain vulnerable populations because it extends this right to minors: if a business is aware that a user is under the age of 16, it must obtain affirmative authorization from the user (ages 13-16) or from a parent or guardian (under 13) before selling or sharing personal information.[74]

Should anonymized data remain exempted?

---

68 Id, at 448.
69 Parks, supra note 45, at 924.
70 Id.
71 Id.
72 Lee, supra note 15, at 152.
73 Carter, supra note 42, at 454.
74 Cal. Civ. Code § 1798.120(c).

Beyond the publicly available information exemption, it is also worth examining the CCPA's deidentified information exemption, which applies to "information that is deidentified or aggregate consumer information."[75] Some data brokers delete identifiers such as names, phone numbers, or email addresses as a "get-out-of-jail-free card," and this is especially true under the CCPA's broad exemption.[76]

However, scientists have demonstrated that a motivated actor can utilize multiple data sets to easily de-anonymize this data.[77] By cross-referencing IMDb's records against the Netflix Prize database, which contained, at the time, the anonymized reviews of 500,000 Netflix subscribers, researchers at UT Austin were able to successfully identify "the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information."[78] That was back in 2007. The *New York Times* has also demonstrated that it is possible to identify an "anonymous dot, by seeing where the device spent nights and using public records to figure out who lived there."[79] From there, it is possible to reconstruct an intimate portrait of that dot's life—places of worship that the dot frequented, a protest that the dot attended, and so on. As mentioned above, anyone from private individuals to law enforcement agencies can purchase this data, deidentified or otherwise, and advances in machine learning will only make deanonymization easier.[80]

For the same reasons that a categorical ban on scraping might prove harmful, a categorical ban on deidentified data would be impossible. However, it may be worth reexamining the types of datasets that people create. Some scholars advocate for viewing "automated mass collection and use of personal data through scraping as a privilege," a framework that would only allow scraping "when it is necessary to further the public interest."[81] As for deciding the scope of what

---

75 Cal. Civ. Code § 1798.140(v)(3).
76 Wang, supra note 24, at 2112.
77 Id.
78 Arvind Narayanan & Vitaly Shmatikov, Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset), (2024), https://arxiv.org/pdf/cs/0610105.
79 Jennifer Valentino-Devries et al., Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret, The New York Times, Dec. 10, 2018, https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html.
80 Wang, supra note 24.
81 Solove, supra note 7, at 1577.

"public interest" entails, there is a wide array of possibilities. It will be important to consider perspectives from researchers, advertisers, journalists, and other groups that rely on web scraping.

Conclusion

The publicly available information exemption of the California Privacy Rights Act enables the automated mass extraction of sensitive information from the Internet. As artificial intelligence algorithms increase in complexity, so too will the demand for this publicly available data, the rate of its collection, and the complexity of potentially sensitive inferences. Consequently, amending this exemption would help to safeguard the privacy rights of California residents from overreach by data brokers, advertising platforms, hackers, and law enforcement agencies. However, legislators must take care to avoid a stringent limit on data scraping, which would likely harm academics, journalists, AI researchers, advertisers, and anybody who uses a search engine. By drawing inspiration from legislation such as the GDPR, remaining mindful of free speech protections for journalists, and taking care not to stifle industry or innovation, policymakers can help the CCPA and CPRA evolve alongside new technological advances.