Via email.

November 8, 2021

California Privacy Protection Agency  
Attn: Debra Castanon  
915 Capitol Mall, Suite 350A  
Sacramento, CA 95814

Re: Invitation for Preliminary Comments on Proposed Rulemaking Under the California Privacy Rights Act of 2020 (PRO 01-21)

I. Introduction

The Center for Democracy & Technology (CDT) respectfully submits these comments in response to the invitation of the California Privacy Protection Agency (“Agency”) for preliminary comments on proposed rulemaking under the California Privacy Rights Act (CPRA). CDT is a nonpartisan, nonprofit 501(c)(3) organization that is dedicated to advancing civil liberties and civil rights in the digital age and challenging exploitative and discriminatory uses of technology. CDT’s focus includes privacy and the responsible use of data and algorithmic decision-making by commercial enterprises and in the administration of government-funded programs and services.

These comments will focus on areas where commercial data practices implicate fundamental rights, including private, for-profit entities that contract with and provide services for governmental entities. Specifically, these comments call on the Agency to help:

- equip consumers to hold automated decision-making systems accountable for bias and denying access to fundamental rights;
- establish sufficient standards for deidentification of data and restrictions on its use;
- ensure appropriate training for staff that use algorithmic systems;
- ensure that businesses’ collection and use of sensitive personal information are subject not only to an opt-out right, but also to additional safeguards that restrict the collection and processing of such information; and
- avoid unintended consequences for businesses that provide services to governmental entities, by ensuring that CPRA regulations appropriately distinguish what rights and duties apply with respect to data collected and processed by such service provider.
II. The Agency should ensure that consumers have access to information about automated and algorithmic decision-making to guard against algorithmic bias and protect their fundamental rights.

The CPRA requires the Agency to promulgate regulations that govern “access and opt-out rights with respect to businesses’ use of automated decision-making technology, including profiling” and that require businesses to respond to consumers’ access requests with “meaningful information about the logic involved in those decision-making processes, as well as a description of the likely outcome of the process with respect to the consumer.”\(^1\) The CPRA also elaborates on what “meaningful information” should entail. The regulations must require businesses to provide notices and information “in a manner that may be easily understood by the average consumer, are accessible to consumers with disabilities, and are available in the language primarily used to interact with the consumer.”\(^2\)

The right to access information about automated decision-making (ADM) should include information necessary for consumers to understand the decision that was made and how it was made. At minimum, the right to access should include the principal reasons for adverse actions, specific data used in the decision, and how the system arrived at its output.\(^3\) Moreover, explanations of data and decisions should be “psychologically coherent,” meaning that the information provided to consumers should be more than a list of variables, but a humanly intelligible explanation of what factors distinguished one decision from another.\(^4\) Further, the explanation should be “faithful” to the system, reflecting how the system actually generated its particular decision.\(^5\)

The CPRA does not define “automated decision-making,” though it does define “profiling” as the “automated processing of personal information ... to analyze or predict aspects concerning [a] natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location, or movements.”\(^6\) Regulations governing access to the “use of automated decision-making technology” must encompass both the algorithm and other technical information and the overall decision-making context in which the technology is used. That is, the regulations should encompass two components: (a) information about the design, training data and methods, logic, input, and output of the algorithm involved in the decision-making process, and (b) the overall decision-

---

5 Id.
In developing regulations, the Agency should ensure that consumers will have access to the information needed to detect the most concerning practices and harms of automated decision-making (ADM) and algorithmic systems, including in particular “where the use of biased AI could raise human rights concerns or violate anti-discrimination laws.” Specifically, the regulation should enable access to information that will reveal disparate impact, as ADM systems often execute decision-making policies in a facially neutral manner that makes it harder to detect discriminatory effects. Three areas in which ADM is increasingly being deployed -- housing, employment, and education -- demonstrate the importance of ensuring the Agency’s regulations provide access to the information needed to determine the existence of discrimination:

8 Id. at 9, 13 (describing how dropout early warning systems have been misused or caused necessary resources to be misdirected).
9 As described below, the CPRA does not define “automated decision-making” and it does not refer to algorithmic decision-making or algorithmic systems. However, an “algorithm” is a “process performed by a computer to answer a question or carry out a task, such as sorting students into schools or classifying social media posts,” and “algorithmic decision-making” is “a decision system that involves algorithms, human decision-makers, legal and social structures, and other forces.” Id. at 6-8. Although “automated decision-making” and “algorithmic decision-making” are not identical, there is substantial overlap between the terms. See European Parliamentary Research Service, Understanding Algorithmic Decision-Making 3-4 (2019), https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf; MARK MCCARTHY, BROOKINGS, FAIRNESS IN ALGORITHMIC DECISION-MAKING (2019), https://www.brookings.edu/research/fairness-in-algorithmic-decision-making. We encourage the Agency to include algorithmic decision-making within the scope of its rulemaking.
● **Housing**: Housing providers use ADM systems that evaluate similar types of data as are used in consumer finance decisions.\(^\text{12}\) This data can include credit, education, employment, and criminal history; income; public records; and banking, purchase, and web activity.\(^\text{13}\) Yet some of this data may be proxies for racism or ableism or lead to disparate impact and inequity in housing for marginalized communities. For example, the ostensible purpose of looking at criminal history is to avoid exposing current residents to new residents who may pose a threat. But the U.S. Department of Housing and Urban Development has advised that blanket prohibitions based on criminal records can be discriminatory.\(^\text{14}\)

Challenging the outcomes of ADM systems used in housing decisions requires access to information about how the applicant’s data was processed through ADM and the extent to which ADM influences the ultimate decision.\(^\text{15}\) Without access to this information, applicants cannot show they were denied housing based on proxies for protected traits, flag risks of disparate impact, or offer additional information that shows why they would in fact be able to meet their obligations should they be approved.\(^\text{16}\) Thus, to address algorithmic bias in housing, it is crucial that consumers are provided access to information about what data is used for decision-making and how ADM processes this data, with a meaningful chance to respond.

● **Hiring**: ADM is also increasingly common in hiring processes and has had disparate impacts on job applicants in many ways.\(^\text{17}\) For example, CDT testified before the California Department of Fair Employment and Housing this year about how various algorithm-driven hiring decisions can worsen hiring disparities for job applicants with disabilities.\(^\text{18}\) Resume parsing tools have rejected applicants whose resumes lack language that the tools were designed to or learned to look for, or that had employment gaps that may be due to disability or extended illness,


\(^{13}\) Comments on Financial Institutions’ AI Use, supra note 3, at 2.


\(^{15}\) Comments to HUD, supra note 11.

\(^{16}\) Id.


\(^{18}\) Testimony of Lydia X.Z. Brown, supra note 17.
pregnancy, or caregiving needs. Gamified aptitude tests, video interview analysis, and personality tests have assessed characteristics or behaviors that often are not relevant to how applicants would be required to perform on the job or would perform if they received accommodations in the workplace. Thus, applicants have been denied job opportunities not because they cannot perform, but because data related to their traits, behaviors, or other factors affected by protected characteristics do not mirror data about “high-performing” employees.

When employers use these types of hiring technologies to assess applicants, they tend not to give applicants advance notice about the manner in which their application materials or they personally will be evaluated, or the criteria based on which the applicants may be disqualified.

The introduction of ADM to the hiring process has also made it less likely for applicants to access information about why they have received an adverse decision. Similar to the housing context, without being provided with such information, they will not be able to challenge discriminatory hiring decisions under laws such as Title VII, the Americans with Disabilities Act, or state employment discrimination law.

- Education: K-12 educational agencies and institutions are navigating a growing market of ADM tools designed to transform a wide range of district and school functions such as assigning students to schools, preventing dropout, and keeping students safe. ADM is also used to scan students’ documents and messages for sexual material and signs of self-harm, bullying, or drug or alcohol use and initiate intervention by administrators or even law enforcement. These decisions can significantly affect students’ experiences, relationships, and future opportunities, whether by determining which school a student attends or by deciding whether or not that student is a threat to school safety.

---

19 Id.
20 Id.
22 42 U.S.C. §2000e et seq.
23 42 U.S.C. §12111 et seq.
24 Cal. Gov’t Code §12940.
25 ALGORITHMIC SYSTEMS IN EDUCATION, supra note 7, at 6-8.
Again, however, the use of ADM can lead to discrimination. For example, schools are increasingly using facial recognition technology, which relies on ADM, for proctoring exams, protecting student safety, monitoring unusual behavior, or even enforcing health and safety measures such as social distancing. Facial recognition technology, however, disproportionately misidentifies students of color, especially Black students, and may further marginalize them by subjecting them to increased interactions with police and school disciplinary systems. Proctoring software struggles to recognize students of color, especially Black students, and disproportionately flags the behavior of students with disabilities, whose movements or accommodations may be flagged by the algorithm as suspicious. The Agency’s regulations should ensure that students and their parents have access to sufficient information about the use of these types of technologies, how they work, how the algorithms were trained, how the ADM tool is used in the overall decision-making process, and other information necessary to determine whether use of the ADM is resulting in bias or discrimination.

### III. The regulations should articulate a high bar for truly deidentifying data and recognize that data harms extend beyond individuals by placing restrictions on the use of deidentified data.

*Responsive to Question 8(c).*

The CPRA and its predecessor, the California Consumer Privacy Act (CCPA), apply only to information that “could reasonably be linked, directly or indirectly, with a particular consumer or household,”

---


30 Id. at 32, 44.


which both laws label “personal information.”

Neither act provides protections for “deidentified” information — or information that cannot be linked to a particular person. Without those protections, supposedly deidentified data can pose risks for both individuals and groups. The Agency should promulgate regulations that take three steps to help ensure that “deidentified” data remains deidentified and to limit secondary uses of even deidentified data.

First, the Agency should ensure that deidentified data stays that way. Reidentification of data has become increasingly feasible as the amount of publicly available data has increased, creating privacy risks for individuals. For example, just four points of “anonymous” location data are enough to uniquely identify individuals 95 percent of the time, and research has demonstrated that health records may be reidentified by cross-referencing publicly available records. Complex datasets with increasing numbers of data points can pose significant obstacles to truly deidentifying data.

The CPRA requires a business to take “reasonable measures” to avoid reidentification, “publicly commit[ting] . . . not to attempt to reidentify the information,” and contractually ensuring that recipients of the deidentified data are bound by the same obligations. Regulations under the CPRA should make explicit that “reasonable measures” include technical safeguards to prevent reidentification of individuals and procedural safeguards, including internal policies that prohibit reidentification.

Second, businesses should be required to describe their methods for deidentifying data in their risk assessments under the CPRA, accompanied by an assessment of the risk of reidentification and the measures taken to mitigate that risk. As the National Institute of Standards and Technology has recognized, “[b]ecause an important goal of de-identification is to prevent unauthorized re-identification, such attempts [at re-identification] are sometimes called re-identification attacks,” and

33 Cal. Civ. Code § 1798.140(o) (effective Jan. 1, 2020); id. § 1798.140(v) (operative Jan. 1, 2023); accord 15 U.S.C. § 6501(b) (defining “personal information” under the Children’s Internet Privacy Protection Act); 34 C.F.R. § 99.3 (defining “personally identifiable information from education records” under the Family Educational Rights and Privacy Act); 45 C.F.R. § 160.103 (defining “individually identifiable health information” under the Health Insurance Portability and Accountability Act)


are akin to cybersecurity risks.\footnote{Simons Garfinkel, Nat’l Institute of Standards & Tech., De-Identification of Personal Information 9-10 (2015), \url{https://csrc.nist.gov/publications/detail/nistir/8053/final}.} To help mitigate those risks, “it is important to understand the techniques and business rules that are being applied when taking steps to remove personally identifiable information” because “depending on the approach, data may still be recoverable.”\footnote{Comments on NIST Proposal, supra note 10, at 2.}

Finally, the regulations should require deidentified data to be accompanied by use and redisclosure limitations. Even data that cannot be reidentified may still pose harms for groups and the people that compose them. Deidentified data may be used to train algorithmic or automated decision-making, which may then perpetuate harms on populations due to biases embedded in the training data.\footnote{Comments on Financial Institutions’ AI Use, supra note 3, at 5-6.} Deidentified data has been used broadly for ADM in critical fields such as housing,\footnote{Relman Colfax, Fair Lending Monitorship of Upstart Network’s Lending Model 18-22 (2021), \url{https://www.relmanlaw.com/media/news/1089_Upstart_Initial_Report_-_Final.pdf}.} credit,\footnote{Algorithmic Systems in Education, supra note 7, at 9.} and education.\footnote{Todd Feathers, Major Universities Are Using Race as a “High Impact Predictor” of Student Success, The Markup (Mar. 2, 2021), \url{https://themarkup.org/news/2021/03/02/major-universities-are-using-race-as-a-high-impact-predictor-of-student-success}.} For example, in education, dropout early warning systems may involve machine learning trained on data that encompasses a broad range of factors like attendance, behavioral information, home and family stability, demographics, and how the student is faring relative to similarly situated students.\footnote{Balancing the Scale of Student Data Deletion and Retention in Education, supra note 41, at 16.} The use of deidentified or aggregate datasets may result in “large disparities in how the software treats students of different races,” which may directly impact students’ educational opportunities.\footnote{Id. at 15.}

Secondary uses of deidentified data may also pose challenges to maintaining public trust in the stewards of the data or ensuring that an individual’s consent is meaningfully respected. Limiting sharing and reuse helps protect against reidentification, harmful secondary uses, and violations of individuals’ original consent.\footnote{Algorithmic Systems in Education, supra note 7, at 9.} Secondary uses may include data that is repurposed and aggregated for research. Algorithmic or automated decision-making systems often rely on repurposed data from disparate, integrated data sets to identify unanticipated patterns, which incentivizes data holders to integrate and repurpose data sets without knowing in advance how the data will be used.\footnote{Id. at 15.} While repurposing data may be useful for gaining insights and improving systems, it complicates other data

---

42 Comments on NIST Proposal, supra note 10, at 2.
44 Comments on Financial Institutions’ AI Use, supra note 3, at 5-6.
48 Balancing the Scale of Student Data Deletion and Retention in Education, supra note 41, at 16.
49 Id. at 15.
ethics issues like transparency, community engagement, and consent.50 Similarly, deidentifying and repurposing education data for commercial purposes may jeopardize public trust.51

IV. The CPRA addresses training to handle consumer inquiries about how consumers may exercise their data rights, but businesses should train employees to also ensure that data is used responsibly. Responsive to Question 9.

The CCPA regulations and the CPRA require businesses to “establish, document, and comply with a training policy” that informs “all individuals responsible for handling consumer inquiries” about businesses’ practices and compliance with the CPRA and about “how to direct consumers to exercise their rights” under the CPRA.52 While employees should be able to respond to inquiries about how businesses are complying with the CPRA and how consumers may exercise agency of their own data, this is not enough. Employees must be equipped to hold businesses accountable for the data practices in which employees may also be engaging.

The CPRA regulations should expand training requirements to educate employees about the ethical use of data.53 Training should inform employees about restrictions on their access to consumer data and on secondary uses of consumer data.54 It should ensure that employees understand the purposes for which data may be disclosed, the necessity of limits on redisclosure, and ramifications for failing to adhere to those limits.55 Deidentification must be carried out only by specific employees with relevant expertise and training in how reidentification of deidentified data can occur, how to ensure sufficient deidentification to reduce the risk of reidentification, how to recognize when data should be thoroughly destroyed, and how to effectively carry out data destruction techniques.56 Finally, changes to data practices may necessitate new or modified training, so these programs should be reviewed frequently and revised to ensure that employees continue to effectively protect consumer data.57

---

50 Id.
51 Benjamin Herold, Schools Collect Tons of Student Information. Deleting It All Is a Major Challenge, EducationWeek (Mar. 15, 2019), https://www.edweek.org/technology/schools-collect-tons-of-student-information-deleting-it-all-is-a-major-challenge/2019/03?cmp=SOC-SHR-FB (“Most vendors don’t really care about data deletion, because they only want to monetize de-identified data, which most policies allow for unlimited use.”).
54 Id. at 14-16.
56 BALANCING THE SCALE OF STUDENT DATA DELETION AND RETENTION IN EDUCATION, supra note 41, at 12-14.
57 DATA SHARING AND PRIVACY DEMANDS IN EDUCATION, supra note 55, at 5-6.
V. The right to opt out of or limit use of sensitive personal information should be accompanied by additional, necessary safeguards.

The CCPA regulation and the CPRA require businesses to provide consumers the choice to opt out of sale and sharing of PI and limit the use and disclosure of sensitive personal information (SPI). Businesses can do so by either providing a link or method on their homepages for consumers to opt out of sharing of PI and use of SPI, or by providing the means to opt out via an opt-out preference signal.\(^{58}\) The CPRA requires regulations to make sure that these options are easy for consumers to use, do not interfere with their online experience, and do not obstruct competition.\(^{59}\) The CPRA regulations must also limit the use of SPI to enable consumers to “exercise their choices without undue burden” and “to prevent business from engaging in deceptive or harassing conduct,” but the CPRA also requires regulations to allow businesses “to inform consumers of the consequences” of opting out of the sale or sharing of PI or of limiting the use of SPI.\(^{60}\)

In addition to providing an opt-out right, the CPRA also requires the Agency to issue regulations, with the goal of strengthening consumer privacy while considering the legitimate operational interests of businesses, to govern the use or disclosure of a consumer’s sensitive personal information, including “[d]etermining any additional purposes for which a business may use or disclose a consumer’s sensitive personal information.”\(^{61}\)

While the right to opt out can help consumers exercise some control over how their data is used and shared, opt-out options put the onus on consumers to protect themselves, which is less effective to protect their rights.\(^{62}\) This burden should belong to businesses. Accordingly, the Agency should impose basic rules that limit a business’s ability to use and disclose particularly sensitive personal information. In particular, regulations must require businesses to put in place ethical use, purpose, and disclosure guardrails to protect consumers’ rights regarding the use of SPI. These protections should include:

- Prohibiting data use that harms individuals or groups;\(^{63}\)
- Require an entity to minimize the data it collects and processes based on the purpose for which the entity needs data (e.g., to provide a product or service requested by a consumer)

\(^{58}\) Cal. Civ. Code §1798.135(a)-(b); Cal. Code Regs. tit. 11, §999.315(a),(c),(f).
\(^{63}\) DATA ETHICS IN EDUCATION AND THE SOCIAL SECTOR, supra note 53, at 5.
• Prohibit unfair data practices, particularly the repurposing or secondary use or sharing of sensitive data without the express, opt-in consent of the consumer;  

• Requiring procedures for determining when data is no longer needed and for completing data destruction;  

• Prescribing procedures for accountability, redress, and mitigation of algorithm-driven disparate impact; and  

• Requiring that the process for developing an opt-out preference signal engages a wide base of stakeholders, including consumer groups, governmental entities that contract with businesses, and technology vendors, who among other considerations can collectively evaluate the merits of selective consent or global opt-out.  

The CPRA also calls for the regulatory process to solicit public participation in “[u]pdating or adding categories of personal information [and] categories of sensitive personal information to those enumerated... in order to address changes in technology, data collection practices, obstacles to implementation, and privacy concerns.” Regulations must be proactive in this area, with additional safeguards to protect a wider range of SPI categories. Under the CPRA, “sensitive personal information” is PI that reveals, among other types of data, social security numbers; financial account and account log-in details; precise geolocation information; information about race, ethnicity, sexual orientation, religious or philosophical beliefs; and genetic data. Some protected classes are included among these types of information, but CPRA regulations must recognize other protected classes as SPI, including gender identity, disability, and immigration status. In addition, some of the currently enumerated types of SPI, such as social security numbers and financial account details, reflect increased risk of financial harm to all consumers, while others reflect data about protected classes that can cause biased decision-making. The regulations should ensure that safeguards for SPI overall are tailored to the different risks involved for each type of SPI.


65 DATA ETHICS IN EDUCATION AND THE SOCIAL SECTOR, supra note 53, at 12-14.  


The CPRA regulations should also revisit how the CPRA covers inferences. In its definition of PI, the CPRA includes “[i]nferences drawn from [other types of PI] to create a profile about a consumer reflecting the consumer’s preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes.” These elements can be proxies for other protected traits, so the resulting “inferences” should be recognized as SPI, not just PI. Therefore, opt-out and other protections that the CPRA provides for SPI should extend to inferences of sensitive data and other proxies for sensitive data, subject to a disparate impact analysis.

The CPRA regulations must limit any exceptions to rights regarding SPI. The CPRA itself already creates an exception allowing the use and sharing of SPI that “is collected or processed without the purpose of inferring characteristics about a consumer,” subject to regulation, so this exception should not be expanded. Even when a data practice is not done with the intention of inferring characteristics about a consumer, the collection, use, or disclosure of sensitive data can still harm individual consumers and protected groups at large. Therefore, exceptions to opt-out rights should only be considered when the business identifies a clear purpose, intended use, and demonstrable need for the data. The business must provide assurances that the data will be subject to the safeguards described above and destroyed when no longer needed, with explicit procedures for redress if these requirements are not met.

VI. Regulations under the CPRA should avoid unintended consequences that would result from requiring service providers for governmental entities to respond to consumer requests under the CPRA.

For-profit and not-for-profit entities provide data processing to support critical governmental services, such as through cloud infrastructure, videoconferencing, web hosting, and supporting remote learning. It is critical that the Agency’s regulations continue to recognize and accommodate the role of service providers for governmental entities and not inadvertently subject the governmental data they hold to rules directed toward private, for-profit entities.

73 Comments on Algorithms and Title IX, supra note 66; Comments on Algorithms and Title VI, supra note 66.
The distinction between a “service provider” and a “business” is well established in privacy law. The distinction is critical to ensuring that there are clear duties among the entities that are ultimately responsible for personal information—a “business” under the CCPA and CPRA—and the entities that they contract with to process the information—a “service provider.”

That distinction is particularly important for service providers for governmental entities. Governmental entities such as schools contract with private businesses to provide services such as cloud storage, student information systems, educational applications, or access to online services. Data held on behalf of governmental entities may be necessary to support governmental services, be particularly sensitive, or be subject to specific laws regarding public access and privacy; consequently, it is important that the responsible governmental entity retains ultimate control over the governmental data held by its contractors.

Recognizing the unique role of service providers for governmental entities, the regulations under the CCPA clarified that a service provider for a nonprofit or governmental entity is not subject to the “full panoply of CCPA obligations,” but rather must collect, use, and destroy data only as directed by the controlling nonprofit or governmental entity. The California Attorney General explained the importance of the rule:

[A] public school district may use a service provider to secure student information, including each student’s grades and disciplinary record. Without this regulation, service providers used by public and nonprofit entities may be required to disclose or delete records in response to consumer requests because they may constitute businesses that maintain consumers’ personal information. Service providers for public and nonprofit entities could also be asked to disclose personal information maintained by a government agency, despite the fact that such files may be expressly exempt from disclosure under the Public Records Act.

---


78 Cal. Code Regs. tit. 11, § 999.314(a) (business that provides services to a non-business must adhere to the “service provider” provisions of the CCPA).

The existing rule mitigates the “unintended result” under the CCPA that governmental data held by a for-profit business might be subject to the CPRA’s rights to access, correct, and delete individual information, despite existing laws governing the disclosure of public records and the privacy of governmental data.

New regulations under the CPRA should maintain the current treatment of service providers for governmental entities. The CPRA reiterated the CCPA’s definition of service providers as acting “on behalf of a business” without addressing the issue of service providers for governmental entities; consequently, it is important that the Agency maintain these vital protections in its regulations. Doing so avoids situations where sensitive data such as a student’s academic performance or accommodations for disabilities would be deleted or altered and helps ensure that governmental entities such as schools can provide services efficiently and effectively. It maintains the balance of the public’s rights to access public records and to privacy in governmental data that has long been established in existing law.

VII. Conclusion

CDT appreciates the Agency’s focus on addressing the impact that businesses’ data practices have on consumers. In advancing the regulatory process, we urge the Agency to prioritize the impact that private entities’ data and ADM practices have for those seeking to exercise fundamental rights. Nondiscrimination and appropriate scoping of obligations, safeguards, and exceptions are vital to ensuring that for-profit data practices avoid data exploitation and serve consumer interests.

Respectfully submitted,

Cody Venzke
Policy Counsel, Equity in Civic Tech Project
Center for Democracy & Technology
cvenzke@cdt.org

Ridhi Shetty
Policy Counsel, Privacy & Data Project
Center for Democracy & Technology
rshetty@cdt.org