Stanford University
Human-Centered
Artificial Intelligence

November 7, 2021

California Privacy Protection Agency
915 Capitol Mall Ste. 350A
Sacramento, CA 95814

Via email: regulations@cppa.ca.gov
Re: PRO 01-21

We are pleased to submit comments in response to the California Privacy Protection Agency's Sept. 22 invitation (Proceeding No. 21-01) for proposed rulemaking under the California Privacy Rights Act of 2020 (CPRA). We are a group of academic researchers and students affiliated with Stanford University and the Stanford Institute for Human-Centered Artificial Intelligence (HAI), and we provide our affiliation for informational purposes only; our comments are made on behalf of ourselves and do not represent the views of either Stanford University or Stanford HAI.

The Agency asked for comment on specific aspects of the CPRA, and we include here several responses as well as additional comments based on our research over the past year on aspects of the CPPA. Our group offers our comments based on our academic expertise and professional experience in the fields of information science (human-computer interaction), computer science (artificial intelligence), and technology policy. The comments are included in the following document.

Thank you for the opportunity to submit comments on these timely and important topics of relevance to all Californians.

Sincerely,

Jennifer King, Ph.D
Data and Privacy Policy Fellow, Stanford Institute for Human-Centered Artificial Intelligence

James Zou, Ph.D
Assistant Professor of Biomedical Data Science and, by courtesy, of Computer Science and of Electrical Engineering, Stanford University

Eli MacKinnon
Graduate Student Researcher, Stanford University

Mitch Bennett
Graduate Student Researcher, Stanford Law School

Catherine Baron
Undergraduate Student Researcher, Stanford University

Divya Nagaraj
Undergraduate Student Researcher, Stanford University

Summary of Recommendations

Topic Two: Automated Decisionmaking

1. We recommend the Agency narrow the scope of covered automated decisionmaking technologies (ADT) to those that relate to a specific outcome of concern, whether it be similar to the GDPR's focus on legal effects, or another interpretation that focuses more directly on outcomes related to consumer privacy.
2. "Profiling" notices should be delivered at a point when consumers can make an actionable decision on whether to submit, and businesses must be incentivized or required to offer substantive alternatives that don't involve the use of profiling.
3. Regarding the provision to consumers of "meaningful information about the logic" of automated decisionmaking processes, we suggest that transparency regarding the data being used to power such processes may be of greater consequence. Giving consumers actionable instructions on how they can prevent such data from being incorporated into automated decisionmaking processes is preferable to focusing exclusively on these processes' logic, which is often hard to interpret even for their designers.
4. Data embedded within machine-learning models must be explicitly covered with respect to consumers' rights to delete, know and correct. This will require regular retrainings of models and potentially the use of novel techniques such as "approximate deletion."

Topic Four: Consumers' Right to Delete, Right to Correct, and Right to Know

5. Businesses should be required to document the source of sensitive personal information they possess on a given consumer, including current contact information for the source parties and whether the information was obtained with explicit and documented consent.
6. In cases where sensitive personal information is not actively needed for exempted business operations and the consumer has not explicitly consented to the collection and use of this information for some other purpose, businesses should be required to permanently delete sensitive personal information by default within a specified time period, even without being requested to do so.

Topic Five: Consumers' Rights to Opt-Out of the Selling or Sharing of Their Personal Information and to Limit the Use and Disclosure of their Sensitive Personal Information

7. CPRA regulations should provide unambiguous language clarifying that global privacy signals, such as the Global Privacy Control (GPC), do not represent merely one among several possible opt-out signals that businesses can choose to recognize, but are instead obligatory to recognize for all businesses.
8. Any browser downloaded by a California consumer, as defined in §1798.40(i), should come with built-in support for GPC and have GPC set on by default.
9. Given potential loopholes, the Agency should require that support for global privacy signals such as GPC be offered in addition to conspicuous opt-out links, not as a replacement sufficient to negate that requirement.
10. Opt-out preferences expressed via one medium (such as a website) should apply automatically to any others (such as an associated mobile app), if it is known from previously collected data that a consumer has expressed such a preference via another medium.
11. Privacy within the mobile app ecosystem, which currently offers no equivalent to the Global Privacy Control, must be prioritized alongside in-browser privacy. The Agency should mandate that apps approach the exercise of CPRA rights in a way that's already been demonstrated to work: a pop-up dialogue displayed upon first use of an app.

Topic Six: Consumers' Rights to Limit the Use and Disclosure of Sensitive Personal Information

12. We recommend that the Agency consider 'precise geolocation' data as a suitable candidate for inclusion in further statutory exemptions from the right of an individual to limit the use and disclosure of SPI; we suggest that precise geolocation data could be collected and processed within a narrow and pre-specified context of use, subject to the limitations we address inline. Further, we recommend that the Agency consider requiring the deletion of this data after delivery of a product or service.

Topic Eight: Definitions and Categories

13. Given the increasing use of AI to attempt to detect or measure individuals' "emotions" and "emotional states," we recommend that these terms comprise their own category of personal information (and potentially, sensitive personal information).

14. The Agency should consider amending the definition of 'sensitive personal information' (SPI) to include inferences that can be characterized as SPI drawn from non-SPI personal information.

15. The Agency should consider amending the definition of "deidentified" to provide further clarity in respect to the reasonableness standard applied to the reidentification risk of anonymized information.

16. We suggest revisions to either the definition of "dark patterns," or to related terms incorporated by reference, in order to allow for a broader interpretation of what constitutes a dark pattern that encompasses novel interfaces, such as voice, that go beyond traditional visual user interfaces.

Topic Nine: Additional Comments

17. Annual Reporting Requirements: CCPA reporting requirements currently produce results that are difficult to collect, compare, and evaluate compliance. We offer several recommendations to improve annual reporting requirements, based on ongoing research by co-author Catherine Baron.

18. Dark Patterns: We provide recommendations to the Agency, via co-author King's recently published work, as to how to further regulation and oversight on this topic beyond consent interfaces.

Topic Two: Automated Decisionmaking

Regarding the use of "profiling" and "automated decisionmaking technology" (ADT), the text of Proposition 24 reads that the Agency will: "Issu[e] regulations governing access and opt-out rights with respect to businesses' use of automated decisionmaking technology, including profiling and requiring businesses' response to access requests to include meaningful information about the logic involved in those decisionmaking processes, as well as a description of the likely outcome of the process with respect to the consumer." Further, "profiling" is defined in the text of Proposition 24 as meaning: "any form of automated processing of personal information, as further defined by regulations pursuant to paragraph (16) of subdivision (a) of Section 1798.185, to evaluate certain personal aspects relating to a natural person and in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location, or movements." This definition appears to have been modified from Recital 71 of the General Data Protection Regulation (GDPR).[1]

We are concerned that the current language of Proposition 24 invites an interpretation that will have the unintended consequence of targeting many algorithmic processes that do not pose inherent privacy risks to consumers. In particular, a key aspect of Recital 71 was omitted from the proposition text: "The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention." Recital 71 incorporates the GDPR's definition of automated processing given in Recital 22, which narrows the applicability of automated processing to decisions which produce *legal effects*.[2] No corresponding definition of automated decisionmaking is included in the proposition text, without which, the applicability of ADT can be broadly interpreted to include any form of ADT whether it presents a privacy risk, or a 'legal effect,' or not.

We recommend the Agency revise this section to narrow the scope of ADT as it relates to a specific outcome of concern, whether it be similar to the GDPR's focus on legal effects, or another interpretation that focuses more directly on outcomes related to consumer privacy. We also recommend that the Agency refine the definition of profiling to focus on the range of processes, automated or not, that contribute to the specific outcomes of concern: the generation of inferences, predictions, and evaluations about individual

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679 (Recital 71).
[2] Ibid (Recital 22).

consumers or groups of consumers conducted without their control, knowledge, or consent. To that end, it bears mentioning that the activity of profiling itself is only one component of this issue; the business practices and technological processes that enable profiling should also be addressed.

We are concerned that an overbroad application of this provision could force the unnecessary labelling of an immensely broad number of ADTs with no privacy risk, providing no useful outcome for California consumers. To the extent that this regulation deliberately or inadvertently targets specific ADTs such as those built using artificial intelligence (AI), and more specifically, machine learning, addressing specific concerns would be better served in separate regulation, where issues related to the data that feeds AI systems can be addressed directly.

We address the specific points that the Agency has asked for guidance on in turn below:

> (a) *What activities should be deemed to constitute "automated decisionmaking technology" and/or "profiling".*

As described above, we are concerned with both the current definitions of ADT and profiling in §1798.185 (16)(a) and this request for comment. At the present level of generality, any algorithmically driven process could be encompassed by this provision, which could have far-reaching negative effects on consumers' online experiences. There are many ADT processes online that have no direct impact on consumer privacy. For example, car rental agencies use ADTs to ask for a consumer's age before displaying rental opportunities. Retail stores may ask for an address or zip code in order to present a list of nearby store locations. In these instances, if the data is used only for the provisioning of the immediate product feature, deeming these types of ADTs as requiring notice through labeling, as well as requiring an explanation of their logic, offers no clear consumer benefit, nor an inherent privacy risk.

It is important to distinguish profiling as a practice of concern distinct from ADT generally. The use of the term "profiling" implies data collection practices, either by a first-party data collector or by one or more third parties, that result in the aggregation of data about an individual[3] that can then be used to classify the individual, group them with other individuals on the basis of one or more characteristics, or make predictions or inferences about them based on past behaviors, actions, preferences, or traits held in common with others. Profiles can be deliberately constructed through the analysis of aggregated data, or emergent, based on identifying correlations between variables without a specific

[3] The data collected through aggregation can be exceedingly diverse, and in addition to specific facts such as demographic data can include mechanisms such as behavioral tracking using browser cookies or third party pixel tracking via web pages, browser fingerprinting, location data, IP addresses, and other similar forms of tracking deployed through mobile apps (both via the apps and third party code embedded within them).

hypothesis (i.e., "data mining"). Profiles can be built through human analysis or through the application of artificial intelligence (e.g., machine learning), and once constructed applied through both manual (human in the loop processes) as well as ADT-based mechanisms. We are concerned that the current definitions of ADT and profiling could exclude profiling practices that do not rely on ADT, or incentivize companies to skirt the regulation by nominally including a human decisionmaker in the process, even if their contribution is minimal.

Given that not all forms of profiling may result in the uncontrolled or adverse collection of personal information, the Agency should consider which practices pose substantial privacy risks, both to individuals as well as groups, or even to society at large. The Agency should also identify the specific practices that enable first and third-party companies to collect and aggregate the data that enable building consumer profiles, in particular those practices that occur without explicit consumer knowledge and consent. For example, some low-level forms of first party website personalization, such as saving a user's preferences, may pose a low privacy risk to consumers as long as the data is collected and used only for this specific purpose and not later sold, shared, or reused outside this context.

Profiling practices that utilize machine learning (ML) bear particular mention here. As we elaborate in further sections below, there are assumptions embedded in the articulations of profiling in the proposition text that are based on conventional non-ML processes, and which may not apply directly to ML-based profiling. ML models are built upon training data—data selected and labeled as representations of specific types of actions or characteristics—that ML algorithms utilize to "learn" and, in the case of consumer profiling, use the results to analyze data and create emergent classification schemes. Instead of a human data scientist analyzing statistical models to identify correlations and create profiles, ML algorithms can create profiles based on the predictions of the trained ML model. This process implicates a different set of challenges for responding to issues of deletion and opting out of them, as we discuss below.

> (b) *When consumers should be able to access information about businesses' use of automated decisionmaking technology and what processes consumers and businesses should follow to facilitate access.*

The timing of any profiling notices should occur at a point where consumers can make an actionable decision on whether to submit, or not, to a profiling-based process, while keeping in mind the many empirical research findings that have demonstrated the challenges with providing effective and meaningful notice.[4] While providing notice may

[4] There is substantial academic literature on this topic, but to offer two overview citations: Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. 2015. A Design Space For Effective Privacy Notices. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security* (*SOUPS '15*). USENIX Association, USA, 1–17; Privacy

imply that consumers have a meaningful *choice* to decline to participate, it does not guarantee that they have a meaningful *alternative*. If providing notice, or labeling, of profiling is intended to follow in the model of notice and consent, then the practice is, unfortunately, futile. Forcing companies to provide notice of profiling if there is no substantive action consumers can take may be confusing at best, frustrating at the very least, and fail to curb use of profiling through public exposure of the practice.

Furthermore, providing notice, or labeling, of profiling may be especially complex given that the creation of profiles themselves likely does not happen in real time when a consumer uses a product. Unless instructed otherwise, companies will bury any notice regarding profiling technologies into their privacy policies, documents that it is well established consumers do not read. As presently written, companies would be incentivized to simply give a notification using the same "take it or leave it" terms that currently exist throughout the online sphere without altering their existing practices, which some have rightly called "consent theatre." If the goal is to curb the excessive or exploitative use of ADTs that undermine privacy, businesses must be incentivized or required to offer substantively the same service without the use of profiling.

> (c) *What information businesses must provide to consumers in response to access requests, including what businesses must do in order to provide "meaningful information about the logic" involved in the automated decisionmaking process.*

It is unclear that providing consumers with meaningful information about the "logic" behind an automated profiling process will be beneficial unless, again, consumers have substantive options to avoid it.

There have been research-driven attempts to provide people with increased transparency around AI and automated decision models.[5] The results are unclear and have raised concerns about the effectiveness of directly passing information, such as model parameters and weights, to users. Even if such information would prove useful to an algorithmically literate individual, it is unclear whether it would substantially impact their actions on a platform. A second large challenge related to algorithmic explainability is that even machine learning engineers, the architects of the very algorithms being analyzed, often cannot interpret the contributions of various weights to the final prediction of the algorithm. Current research aims to improve the interpretability of key models in specific

and Human Behavior in the Age of Information, Acquisti, Alessandro and Brandimarte, Laura and Loewenstein, George. *Science*, *347* (*6221*), *509--514*, 2015.

[5] For example, see: Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods.Entropy 2021, 23, 18. https://dx.doi.org/10.3390/e23010018; Katherine Miller. Should AI Models be Explainable? That Depends. March 16, 2021, available at: https://hai.stanford.edu/news/should-ai-models-be-explainable-depends.

domains—like facial recognition—but at present, there are no guarantees that any gains in interpretability resulting from this research will generalize to other models and systems.

Given these significant challenges around algorithmic interpretation, we suggest that what may be of greater consequence is transparency regarding the *data being used to power such processes*, with clear instructions to consumers as to how they can prevent such data from being incorporated into profiling processes, rather than focusing exclusively on the logic of the process itself.

We ask the Agency to consider requiring companies to document the source of all the data they collect, purchase, or trade, including documentation regarding whether or how the consumer was asked to consent to the collection of the data. We conjecture that if companies had to document the provenance of the data they collect, and present this information to regulators and to consumers, then consumers might be able to draw meaningful conclusions from companies' use of it. For example, if information brokers had to reveal all sources of all data collected about an individual (including the contact information for the source parties), a consumer armed with such detail might be able to trace the origins of particular data points, including incorrect or outdated information. To an extent, credit reporting bureaus are required to engage in a form of this practice today. Within the context of privacy, this may be a more powerful and relevant approach than requiring transparency of ADT logic alone.

> (d) *The scope of consumers' opt-out rights with regard to automated decisionmaking, and what processes consumers and businesses should follow to facilitate opt outs.*

If labeling profiling is to be effective, then consumers need to be given actionable options to refuse it, as well as reasonable and realistic alternatives to the profiling-based service. The take-it-or-leave-it terms that consumers are offered today force many to engage with companies or with business practices that they would otherwise prefer not to, given that in many cases they lack other options. However, the Agency's determination of what constitutes legitimate profiling will matter here, as presumably there are some products or services for which it would be difficult to offer a viable alternative. We offer more specifics regarding the challenges and consequences of opting out of ADTs based on artificial intelligence in our comments on Topic 4 below.

Topic Four: Consumers' Right to Delete, Right to Correct, and Right to Know

Regarding consumers' rights to correct, delete and know what personal data a business has collected, we alert the Agency to the need to plan for the subtleties that such rights will

entail when applied to personal data used to train machine learning (ML) models. While it is easy to imagine some piece of personal information as an inert entity in a data table — easily deleted or corrected — the reality is that a business can quickly propagate information through the tools it uses in ways that create obstacles to straightforward removal and alteration. In particular, data used to train ML models becomes "embedded" in those models in ways that are not easily reversed. In spite of the relative difficulty posed by amending the data that undergirds actively deployed ML models, privacy rights must extend to this data in order to be meaningful; in exactly these contexts, data carries far-reaching impacts and the potential for unwanted distribution and disclosure.

ML models are first trained on one dataset before being applied to the analysis of novel data. For example, an employment screening business might collect information on a broad range of individual characteristics, such as age, geolocation, educational background and past purchasing behavior, before using this data to train a filtering algorithm for job applicants. In the training phase, the algorithm will surface correlations between individuals' specific personal characteristics and their success as job applicants with respect to some role. Then, when the training is complete, this algorithm will be used to infer the suitability of new applicants—classifying them based on how well they match the patterns embedded in the training data. Imagine that a California consumer whose data was used to train the model requests this business delete their data. At this stage, even if their individual record were deleted, their contribution to the ML model would remain intact until the model is retrained on an updated data set.

As long as the model remains in use, any erroneous, outdated or simply unwanted correlations that an individual's data contributed to will continue to manifest and subvert the relevant individual's rights over their personal information. In fact, failure to remove data from ML models would directly nullify a consumer's ability to meaningfully control the use and potential spread of their personal information: Researchers have shown that under some conditions, original training data can be reconstructed and ultimately deanonymized by analyzing the behavior of an ML model that incorporates it.[6] While an individual's data remains embedded in a model, it cannot be said to have been deleted. The Federal Trade Commission supported this view in a recent settlement with a photo-sharing app that allegedly deceived consumers about how it was applying facial-recognition technology — the settlement required that all models and algorithms trained using the data be deleted along with the original photo data.[7]

[6] Salem, et. al. *ML-Leaks: Model and Data Independent Membership Inference Attacks and Defenses on Machine Learning Models* (February 2019).
[7] *Everalbum, Inc., In the Matter.* Case summary and decision available at: https://www.ftc.gov/enforcement/cases-proceedings/192-3172/everalbum-inc-matter.

If a consumer requests deletion or alteration of their data used in an ML model, the most straightforward way for a business to honor the request is to retrain every ML model in which these data were included. Retraining an ML model is not a trivial process and can be both time-consuming and expensive, given the computational resources required to analyze vast datasets. For this reason, we anticipate that the Agency will see significant pushback from businesses wishing to avoid such a responsibility and potential cost center. This pushback may include arguments proposing, in effect, that deleting or altering data already embedded within an ML model is onerous or virtually impossible.

The Agency should reject such arguments and ensure that data embedded within ML models are explicitly covered with respect to consumers' rights to delete, know and correct. While retraining models is time- and resource-intensive, businesses of the size and specialization covered by the CPRA already routinely retrain models to improve them as new data are collected. Additional retrainings for the purposes of honoring CPRA requests are both feasible and necessary to honor the law's intent, and any subsequent regulation should be written with this requirement in mind, particularly with regards to timing requirements. Moreover, the understanding that data included in ML models is subject to deletion and alteration requests will incentivize businesses to be both more conservative in their collection and use of personal data, and more explicit in communicating to consumers which data they use and how they use it, as well as in obtaining consent—businesses will be motivated to avoid mandatory retraining or penalties resulting from the misuse of individuals' personal information.

In addition to retraining ML models on new data, there are other avenues for promptly honoring consumer data rights. A team of researchers from UC San Diego and Stanford University, including Professor James Zou, a co-author of this comment, has advanced a technique called "approximate data deletion."[8] Using this technique, the impact of specified data on an ML model can be quickly and cheaply negated, so that the potential for deducing these data in their raw form is greatly reduced or eliminated. The application of this method or a similar one would also allow businesses to respond to user requests immediately and without taking a model offline during retraining — it therefore might form a stopgap that could be used by businesses to honor consumer data rights before they've had an opportunity to fully retrain a model with relevant data deleted.

[8] Izzo, Smart, Chaudhuri and Zou. *Approximate Data Deletion from Machine Learning Models* (April 2021). Available at: http://proceedings.mlr.press/v130/izzo21a/izzo21a.pdf.

Topic Five: Consumers' Rights to Opt-Out of the Selling or Sharing of Their Personal Information and to Limit the Use and Disclosure of their Sensitive Personal Information

a. *What rules and procedures should be established to allow consumers to limit businesses' use of their sensitive personal information.*

As previously discussed in regard to Topic Two above, we believe consumers' control of their personal information must be rooted in transparency around the information's provenance—this is especially true in the case of sensitive personal information. Businesses should be required to document the source of sensitive personal information they possess on a given consumer, including current contact information for the source parties and whether the information was obtained with explicit and documented consent. Businesses should supply this information to consumers, at minimum in their data privacy disclosures. While not a substitute for an outright ban on non-consensual collection, such disclosures could help empower consumers not only to limit a first-party business's use and disclosure of their sensitive personal information to those exempted purposes explicitly outlined in CPRA, but also to identify specific third-party sources, and, if they wish, take steps to limit its continued spread from those sources as well. A measure such as this would help close one of the existing loopholes in the CCPA: that even with "do not sell" and deletion rights, consumers often have no idea to whom to make these requests beyond the businesses with whom they have first-party relationships.

We also advise that, in cases where sensitive personal information is not actively needed for exempted business operations and the consumer has not explicitly consented to the collection and use of this information for some other purpose, businesses should be required to permanently delete sensitive personal information by default within a specified time period, even without being requested to do so. Barring such a provision, we expect that the CPRA's broad language exempting the collection and use of sensitive personal data for specific purposes (e.g. ensuring "security and integrity") will be abused by businesses to hoard sensitive personal information without meaningful justification.

b. What requirements and technical specifications should define an opt-out preference signal sent by a platform, technology, or mechanism, to indicate a consumer's intent to opt out of the sale or sharing of the consumer's personal information or to limit the use or disclosure of the consumer's sensitive personal information.

As the Agency is clearly aware, based on the important addition to the CPRA of new language prohibiting the use of dark patterns in certain contexts (see "Topic Eight: Definitions and Categories" below for additional discussion), there is a pressing need to standardize the processes by which consumers express their preferences regarding the sale and sharing of their personal information, as well as the use and disclosure of their sensitive personal information. Presently, opt-out preferences are inconsistently designed, can be hard to find, and provide opportunity for consumer manipulation. Best Buy customers, for example, if they manage to locate the "Do Not Sell My Personal Information" link in small print at the bottom of the electronics retailer's homepage, will be greeted with a lecture on the technical definition of the word "sale" before seeing instructions on how to opt out.[9] Though this explanatory text makes mention of a "Do Not Sell My Personal Information" button, a mention which itself could easily be made an interactive link, the button is located further down the page and relegated to the left margin.

Such unnecessary friction is typical, but there is a deeper problem. Even after a consumer completes an opt-out process, that opt-out signal is only valid on a specific device and on a specific browser whose cookies have remained unaltered since the point in time when the opt-out signal was registered. Privacy-minded consumers are perhaps especially likely to regularly delete their browser cookies, negating past opt-out requests in the process, and businesses are therefore incentivized to wage a war of attrition on consumer data preferences. While appearing to honor consumer preferences around their personal information, businesses need only wait for consumers to switch to a new device, a new platform, such as an app, or a new (or just newly reset) browser — perhaps at a time when they are in too much of a hurry to initiate a new opt-out process — before they can safely resume data sales and sharing.

This is why an automatic opt-out mechanism like the Global Privacy Control (GPC) is crucial for supporting consumer privacy in our present data ecosystem—it allows consumers to efficiently and persistently communicate opt-out signals, and to defend against businesses that would try to exploit the ephemerality of manual opt-out requests to subvert their privacy preferences. The CPRA includes language in §1798.135(b)(1) describing opt-out signals sent via "a platform, technology, or mechanism," such as the GPC, and as former Attorney General Becerra clarified last summer, "under law, [GPC] must be honored by covered businesses as a valid consumer request to stop the sale of personal information."[10] However, CPRA regulations should go further and provide unambiguous language clarifying that the GPC is not merely one among several possible opt-out preferences that businesses can choose to recognize, but an opt-out signal that is

[9] BestBuy.com. Accessed Nov. 2021 via: https://www.bestbuy.com/site/california-privacy-rights/do-not-sell/pcmcat1576178819013.c?id=pcmcat1576178819013

[10] "CCPA Frequently Asked Questions." State of California Department of Justice. Accessed Nov. 2021 via: https://oag.ca.gov/privacy/ccpa

obligatory to recognize for any business that is technically capable of doing so (in effect, any business, excluding those that fall under the requirements of the CPRA despite not having a website). Such an addition would help resolve the apparent inconsistency between §1798.135(b)(1) and §1798.135(e), the former of which seems to position global preference signals like GPC as one CPRA-compliant option and the latter of which says that businesses must honor global opt-out signals in all cases. Universal recognition of the GPC will ensure that it empowers consumers to exercise their data preferences in a sustainable manner within a current landscape of inconsistent, often inconspicuous and (thanks to frequent changes in consumers' browser cookies and preferred devices) ephemeral opt-out request processes.

It's important, though, that GPC is not only universally recognized but also universally available. Though GPC is gaining traction, it's currently not supported by either of the U.S.' two most popular browsers—Chrome and Safari, which together account for 84.62% of installed browsers in the U.S.[11] In fact, only one of the country's nine most popular browsers supports it—Firefox, whose share of U.S. browsers is just 3.53%. Given GPC's crucial utility in realizing CPRA's aims, as well as its extreme simplicity and ease of implementation, we recommend that the Agency require any browser downloaded by a California consumer, as defined in §1798.40(i), come with built-in support for GPC and have GPC set on by default.

The GPC's current lack of wide availability creates other threats to CPRA's ultimate effectiveness. As discussed above, the CPRA could currently be interpreted to position the GPC and other similar tools as *alternatives* to the standard opt-out process defined by the CCPA—namely, a conspicuously placed link or set of links on a business's homepage. Section 1798.135(b)(1) reads, in part: "A business shall not be required to comply with subdivision (a) if the business allows consumers to opt out of the sale or sharing of their personal information and to limit the use of their sensitive personal information through an opt-out preference signal sent with the consumer's consent by a platform, technology, or mechanism, based on technical specifications set forth in regulations adopted pursuant to paragraph (20) of subdivision (a) of Section 1798.185…" The aforementioned subdivision (a), with which a GPC-recognizing business is not required to comply, specifies the need for a link or pair of links that consumers can use to express a preference signal with respect to personal information and sensitive personal information.

For a consumer who was not using a GPC-enabled browser or another equivalent mechanism when interacting with a business that had opted to support GPC and exempt itself from the requirements of Section 1798.135(a), there would not necessarily be any other clear method by which to express an opt-out signal. Given that, as detailed above,

[11] "Browser Market Share United States of America." StatCounter Global Stats. Accessed Nov. 2021 via: https://gs.statcounter.com/browser-market-share/all/united-states-of-america

the most popular browsers don't currently support GPC, companies might choose to support GPC as a means to avoid offering a more broadly accessible opt-out signal, such as a conspicuous link. We therefore recommend the Agency require that GPC be supported *in addition* to conspicuous links, not as a replacement sufficient to negate that requirement, in order to close any potential loopholes for consumers.[12]

Consumers currently face another obstacle in expressing their opt-out preferences. Revisiting Best Buy's "Do not sell my information" page, we see the retailer's concise statement of this issue: "Your selection won't cross from the website to mobile application (or the other way round): If you click the "Do Not Sell My Personal Information" on this website, your selection will not transfer to our mobile app. Similarly, your activation of the feature on the mobile app won't apply to bestbuy.com. You'll need to do both."[13]

There are cases in which such siloing of preferences between browsers and apps is unavoidable. For one, it's possible that a business whose website is visited by a consumer who also uses its mobile app has simply not identified them as the same person. However, it's very common that businesses *do* positively match the identity of a website visitor with an app user, via a login, a unique device ID or some other fingerprinting mechanism, though these comments should not be read to promote the use of such practices for the purposes of profiling. In such cases, there is no technical reason why a request submitted via one medium could not be automatically applied to the other, and in terms of consumer preferences, there is no reason why it should not be. After all, behind the distinct access points, it is the same business with the same data, the same data practices and the same incentives. We therefore recommend that the Agency require opt-out preferences expressed via one medium (such as a website) to apply automatically to any others (such as an associated mobile app), if it is known from previously collected data that a consumer has expressed such a preference via another medium.

Mobile apps present other important challenges with regard to opt-out preferences. Though the GPC will go a long way toward empowering consumers to efficiently exercise their data rights on the web, and to aid them in navigating an often daunting variety of opt-out-request formats, it unfortunately does nothing to streamline user preferences signals in the app ecosystem, an equally important domain of data collection. Increasingly, companies take aggressive tactics to encourage consumers visiting their websites via a browser to instead download a proprietary app. Reddit, for example, one of the world's twenty most-visited websites, follows visitors across its website with a floating banner that reads "This page looks better in the app" and includes a download button. For a significant

[12] If the agency accepts our recommendation to require that any browser downloaded by California consumers natively support GPC, then this requirement would be less urgent. However, conspicuous links serve a valuable educational function in all cases.

[13] BestBuy.com. Accessed Nov. 2021 via: https://www.bestbuy.com/site/california-privacy-rights/do-not-sell/pcmcat1576178819013.c?id=pcmcat1576178819013

range of businesses, accessing services via an app is the norm, and this trend will only accelerate, given businesses' myriad incentives to coax consumers into the contained environs of a proprietary app, as well as the genuine conveniences these apps offer. The Agency must therefore ensure that consumers are equally empowered to protect their data rights within the app ecosystem as they are on the open web. To this end, the Agency should require that mobile apps solicit *opt-in consent* for the sharing or sale of personal information upon first use of the app.

Recent changes to how Apple's iOS solicits consumer tracking preferences are instructive with respect to the efficacy of an opt-in system. In Spring of this year, Apple began requiring iOS apps to solicit user consent for allowing the app "to track your activity across other companies' apps and websites" via a unique identifier associated with their device. A global device setting also allows iOS users to reject all of these requests by default. The reception of the feature has been illuminating: When presented with the option, 96% of iOS users chose not to allow cross-service tracking.[14] Clearly, consumers choose privacy when given an accessible, easy-to-interpret choice. (However, recent research also suggests the limits of this feature; even when a user has opted-out of tracking, apps have been able to work around this limitation to continue to track individuals.[15]) There's an oft-overlooked detail to this story, though. The global setting that allows iOS users to opt out of all cross-service tracking, accessible via the phone's settings menu, was available long before the software update that made it mandatory for apps to display opt-in forms within their apps. In other words, the massive surge in tracking opt-outs that followed the update was not due to a new capability on the part of consumers, but rather to a new presentation of that capability via the user interface.

Though consumer response to the new iOS tracking opt-out is a valuable reference point, the feature is no substitute for a "Do not sell or share my personal information" request as envisioned under the CPRA. iOS' built-in feature prevents cross-service tracking, but it does nothing to limit data collection *within* an app or that data's subsequent use. "Do not sell" requests directed at mobile apps must currently be navigated by consumers without help from the mobile platform provider, and they are often just as hard to find within apps as they are on the web, buried in settings pages or at the end of a labyrinth of links. Until all mobile platforms are required to introduce a mobile analogue to the GPC—a setting that would allow consumers to automatically opt out of the sharing and sale of their personal information gathered directly by apps — the Agency should mandate that companies

[14] Axon, Samuel. "96% of US users opt out of app tracking in iOS 14.5, analytics find." https://arstechnica.com/gadgets/2021/05/96-of-us-users-opt-out-of-app-tracking-in-ios-14-5-analytics-find/
[15] Geoffrey Fowler and Tatum Hunter. "When you 'Ask app not to track,' some iPhone apps keep snooping anyway". Washington Post, Sept. 23, 2021. Available at: https://www-washingtonpost-com/technology/2021/09/23/iphone-tracking/ology/2021/09/23/iphone-tracking/.

approach the exercise of consumer data rights in a way that's already been demonstrated to work: a pop-up dialogue displayed upon first use of the app. We recommend this form give consumers the right to *opt in* to the sale and sharing of personal information as opposed to opting out, or at minimum, that it does not pre-select, highlight or otherwise give biased placement to the option to consent to personal information sales and sharing. Apple's recent experimentation in strengthening user choices around privacy has shown that this method of soliciting consent is effective and that consumers are eager to exercise their rights in this way. The Agency should require businesses to adopt this simple, powerful approach and ensure that consumers are fully empowered to make their own decisions on the sale and sharing of their data when accessing services via apps.

c. What technical specifications should be established for an opt-out preference signal that allows the consumer, or the consumer's parent or guardian, to specify that the consumer is less than 13 years of age or at least 13 years of age and less than 16 years of age.

If the Agency accepts our recommendation that any browser downloaded by a California resident both support GPC and have GPC turned on by default, then the sale and sharing of information can only be initiated when a user elects to turn GPC off. In that circumstance, consumers can be presented on a per-business basis with a set of user interface options informing them that: 1) users aged 12 and under may opt-in to selling/sharing only with the affirmative authorization of a parent or guardian; and 2) that users aged 13 and over may opt in directly. We think this would be an improvement over the current set of regulations, which are overly complex, and allow businesses to take advantage of the fact that if a website or app visitor is not known by the business to be under the age of 16, then the business could simply collect information from that visitor as if they were an adult. However, any opt-in user interface elements must be compliant with respect to the dark patterns provisions of the CCPA and the CPRA.

e. What businesses should do to provide consumers who have previously expressed an opt-out preference via an opt-out preference signal with the opportunity to consent to the sale or sharing of their personal information or the use and disclosure of their sensitive personal information.

After the twelve-month opt-out window has passed, a business may ask a consumer directly whether they wish to opt-in to information sale or sharing. We recommend that the Agency provide clear user interface guidelines that demonstrate appropriate methods for initiating this dialogue that prohibit the use of dark patterns (as already articulated regarding consent in the CPRA) or any other design element or language that is deceptive, manipulative, or coercive.

Topic Six: Consumers' Rights to Limit the Use and Disclosure of Sensitive Personal Information

The following two questions on this topic are addressed, in turn, below:

> *a. What constitutes "sensitive personal information" that should be deemed "collected or processed without the purpose of inferring characteristics about a consumer" and therefore not subject to the right to limit use and disclosure.*
>
> *b. What use or disclosure of a consumer's sensitive personal information by businesses should be permissible notwithstanding the consumer's direction to limit the use or disclosure of the consumer's sensitive personal information.*

Our interpretation of the two questions above are as follows: a) Are there any contexts in which SPI should be able to be collected and used without being subject to limits on use and disclosure?; and b) Are there any uses or disclosures by businesses that should be allowable regardless of a consumer's expressed preference to limit their use of SPI? In sum, we suggest that precise geolocation data could be collected and processed within a narrow and pre-specified context of use, subject to the limitations we address below.

The section referenced by the relevant footnote to Topic 6(a) is Civ. Code §1798.121(d) which provides for an exception to the opt-out regime for sensitive personal information (SPI).[16] The existing permissible uses of SPI collected from a consumer that are allowed following a consumers exercise of the opt-out right under subsections (a) and (b) are those:

(1) necessary to perform the services or provide the goods reasonably expected by an average consumer who requests those goods or services;
(2) to perform a limited set of "business purposes" set out in §1798.140(e), namely ensuring the security and integrity of consumer personal information, short-term transient use such as non-personalised advertising (if not disclosed to third parties), operational purposes such as order fulfilment and processing of payments, and quality and safety assurance for services or devices used by the business; and
(3) as otherwise authorised by regulations enacted under §1798.185(19)(C).

Subsection (19)(C)(iv) of §1798.185, to which §1798.121(d) refers, provides detail in respect to the purpose of any further regulations providing for additional categories of exempted SPI. When read together, §1798.121(d) and §1798.185(19)(C) contemplate the

[16] §1798.121(d) provides that '[s]ensitive personal information that is collected or processed without the purpose of inferring characteristics about a consumer is not subject to this section, as further defined in regulations adopted pursuant to subparagraph (C) of paragraph (19) of subdivision (a) of Section 1798.185, and shall be treated as personal information for purposes of all other sections of this act, including Section 1798.100".

making of regulations that identify SPI and contexts of use which is not collected for the purposes of inferring characteristics about a consumer and therefore should be added to those limited permissible uses which apply to SPI even after the relevant consumer has exercised their opt-out right.[17] This is provided that any such permitted uses balance business need and consumer privacy and do not provide a means to circumvent the opt-out protections.

The CPRA introduced the category of sensitive personal information (SPI), we presume, based on the assumption that there are types of information that, irrespective of the context for which it was collected, present a moderate to high risk to individuals should it be disclosed without permission, lost in a breach, or sold to a third party.

Our comments primarily relate to 'precise geolocation' data which has high operational value to businesses providing a number of services and whose use in many contexts is not intended to infer characteristics about the relevant consumer to which that SPI relates. Arguably, there are other types of data, such as one's sequenced personal genomic data, that may carry a similarly high privacy risk and threat of identification or inference, though we limit our discussion here to precise geolocation. Accordingly, we recommend that the Agency consider 'precise geolocation' data as a suitable candidate for inclusion in further statutory exemptions from the right of an individual to limit the use and disclosure of SPI.

While §1798.121(a) currently requires that SPI that is subject to an opt-out request be used only as necessary to perform the relevant service or goods reasonably expected by an average consumer, we believe this limitation is overly broad in the context of 'precise geolocation' data and any permitted uses of such data notwithstanding an opt-out request should be more narrowly tailored. In particular, it is arguable that a 'service' provided to a consumer for which 'precise geolocation' data is allowed to be used may be construed broadly to include a range of ancillary or incidental uses related to the primary purpose for which such information was collected. This is particularly true where an individual profile or account that includes 'precise geolocation' data is applied across a suite of digital services provided by a business.

We continue to emphasize that the risk of inferring characteristics of a consumer based on 'precise geolocation' data remains high and any regulations which contemplate permissible uses of 'precise geolocation' data should address the risk of inferring further

[17] §1798.185(19)(C)(iv) provides that the Agency shall issues regulations with the goal of strengthening consumer privacy while allowing for legitimate operational interests of businesses, including regulations "[e]nsuring that the exemption in subdivision (d) of Section 1798.121 for sensitive personal information applies to information that is collected or processed incidentally, or without the purpose of inferring characteristics about a consumer, while ensuring that businesses do not use the exemption for the purpose of evading consumers' rights to limit the use and disclosure of their sensitive personal information under Section 1798.121".

SPI from use of 'precise geolocation' data, for instance: geolocation attributable to an individual at places of worship, healthcare facilities, or politically meaningful locations or events.

In order for the collection and processing of geolocation data to justifiably be exempt from the broader SPI use/disclosure opt-out limitations, the collection or processing should be: time- and event-limited (i.e. not rolling, aggregated or historical), as well as compliant with any existing data minimization requirements contained within the statute; disclosure to third parties should be prohibited without additional consent regardless of an average consumer's reasonable expectation that such a disclosure might occur; and, individuals should only be locatable at a general (coarse) level of precision.

More broadly, we recommend the Agency consider further regulations in the form of positive obligations on organizations to delete SPI, particularly 'precise geolocation' data, following its time- or event-limited use. Where certain SPI is exempt from the use/disclosure opt-out regime for specific time and event limited purposes, the data minimization obligations in respect to that data should be broader and more onerous. Further regulation may also consider limitations on the cross-referencing of 'precise geolocation' data and biometric identifiers where the risk of attribution to an individual is higher, for instance biometric authentication for payment processing.

Topic Eight: Definitions and Categories

Comment on select questions surrounding the possible update to CCPA- and CPRA-related terms and categories are provided below:

> *a. Updates or additions, if any, that should be made to the categories of "personal information" given in the law.*

"Emotions" or "Emotional state": While §1798.140 (v)(K) includes "preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes" as part of the definition of inferences, given the increasing use of AI to attempt to detect or measure individuals' emotions or emotional state, we recommend that these terms comprise their own category of personal information (and potentially, sensitive personal information).

> *b. Updates or additions, if any, that should be made to the categories of "sensitive personal information" given in the law.*

The Agency should consider amending the definition of 'sensitive personal information' (SPI) to address inferences that can be characterized as SPI drawn from non-SPI personal information. These inferences are of concern when they result from the application of data analytics to personal information relating to an individual for the purposes of generating further salable or marketable insights. A revised definition of SPI should contemplate insights which themselves concern or infer a category of SPI about an individual, whether true or not.

c. Updates, if any, to the law's definitions of "deidentified" and/or "unique identifier."

The Agency should consider amending the definition of 'deidentified' to provide further clarity in respect to the reasonableness standard applied to the reidentification risk of anonymized information. The standard should contemplate a rapidly evolving technological and computing environment and, in respect to SPI, require a standard of care commensurate with the inability of an individual to protect themselves against unauthorized disclosure or misuse stemming from reidentification. The Agency should consider imposing a higher standard of care within the definition of 'deidentified' and/or minimum standards and technical guidance on compliant anonymizing treatments.

j. The regulations, if any, that should be adopted to further define "dark patterns."

We suggest revisions to either the definition of dark patterns, or to related terms incorporated by reference in order to allow for a broader interpretation of what constitutes a dark pattern. Our concern stems from the fact that the present focus of dark patterns research and taxonomy creation has been with static visual user interfaces. However, emergent technologies may also deploy dark patterns, such as voice activated systems, or other new user-computer interfaces that don't fit the category of traditional static visual user interfaces. Additionally, there are open questions about how to classify dynamic or adaptable user interface mechanisms, such as algorithmically driven content feeds, that foster coercive or manipulative digital interactions that again are not easily described as "user interfaces." Broadening the definition of what constitutes an "interface" would ensure that the regulation can adapt to changes that expand past the traditional graphical user interface. We discuss this topic in greater depth in the paper we reference in the following section on dark patterns.

Topic Nine: Additional Comments

We appreciate the opportunity to provide the Agency with additional comments on the following issues which we think are highly relevant to the scope of the CPRA regulations:

1. Revise the CCPA reporting requirements. The observations in this section are based on ongoing research of the CCPA metrics from 100 companies across industries and sizes.

- Overall recommendations:
    - Currently, the scale of CCPA reporting metrics are inconsistent across firms, and this makes comparisons difficult to interpret. Some companies elect to expand CCPA rights to U.S. and global user bases, and their CCPA metric reporting include non-Californians. It would be helpful for research purposes if firms explicitly indicated the scope of their implementation of CCPA in their metric reports, or limited their reporting to Californians only.
    - Inconsistencies in reporting the mean versus median response rates to requests make it difficult to compare performance across companies. Having data on both the mean and median will help contextualize the metrics.
    - Some companies are unclear about the timeframe of the reporting, electing to display metrics, for example, from a few months or splitting across two calendar years.
    - Finally, companies subject to reporting requests should be obligated to submit these metrics to the Agency directly, for the Agency to post publicly and track.
- Access requests:
    - What counts as compliance with access requests is unclear and may vary across companies. Some firms distinguish in their metrics requests where they have provided personal information, as opposed to data categories.
    - Relatedly, most firms do not specify the type of identifiers used to fulfill access requests, which presents a challenge in interpreting the metrics. More specificity along the two dimensions mentioned will enable researchers to better evaluate CCPA's impact.
- Opt-out requests:
    - Companies have a wide interpretation of what constitutes an opt-out request, making it hard to evaluate corporate compliance using these metrics. First, about one-third of the companies in our preliminary study did not disclose metrics related to opt-out requests. The rationale was that they did not sell the personal information of customers. Second, among the companies that did disclose said metrics, many of them interpreted the opt-out requests pursuant to CCPA as equivalent to users' responses to cookie consent management banners. More research is needed to see whether companies treat cookie banners as substitutes for explicit DNS links.
        - The decision to equate cookie consent preferences and DNS preferences tends to lead to impressive response rates, given cookie

consent forms' scale and automation, and purportedly expands the right to opt-out to all users that visit a website. However, this raises the question of whether corporate responses to consumer cookie preferences sufficiently uphold the right to opt-out under CCPA. The problem is that an individual can opt-out multiple times, and, in cases where their browsers clear cookies after a session, they have to re-assert their choices. This not only leads to double counting in CCPA metrics, but also casts doubt on the efficacy of using responses to cookie consent preferences as a measure for opt-out metrics.

- Equating cookie consents and DNS preferences further does not address how consumers may opt out from the sale of personal information when interacting with firms that enable third-party companies to collect, use and share users' personal information, as defined by CCPA. Furthermore, opting out from a website may not automatically translate into opting out from personal information collected in mobile applications.

2. Dark patterns: The CCPA introduced language targeting specific forms of dark patterns observed in CCPA "Do Not Sell" opt-out requests, while the CPRA includes both a definition of "dark patterns," as well as prohibitions focused narrowly on the use of dark patterns in consent mechanisms related to the disclosure of personal information. Co-Author King argued in a recent paper, "*Regulating Privacy Dark Patterns in Practice—Drawing Inspiration from the California Privacy Rights Act,*"[18] that the current language included in the CPRA presents a model for other states and regulatory agencies to follow. However, much of the possibility that the CPRA offers will be determined by how the Agency intends to regulate this area, and whether it does so expansively or conservatively. We offer the commentary in this article as a reference to the Agency on how to approach further regulation of this topic, and note specifically: "[t]he optimal outcome is not one where consumers are given more checkboxes to check and buttons to click in the name of "compliance." If we are not careful about how we interpret coercion and manipulation, consent mechanisms will merely be fragmented into more rote and meaningless actions rather than transformed into new mechanisms that are more substantive, meaningful, and informative. In prohibiting dark patterns, the CPRA creates an opportunity for California to lead by example and develop standards that demonstrate best practices—or light patterns—for consent."[19]

[18] 5 GEO. L. TECH. REV. 250 (2021). Available at: https://georgetownlawtechreview.org/regulating-privacy-dark-patterns-in-practice-drawing-inspiration-from-california-privacy-rights-act/GLTR-09-2021/.
[19] Ibid, at 272.