SENATE JUDICIARY COMMITTEE
Senator Thomas Umberg, Chair
2023-2024 Regular Session

AB 2877 (Bauer-Kahan)
Version: June 17, 2024
Hearing Date: June 25, 2024
Fiscal: Yes
Urgency: No
CK

SUBJECT

California Consumer Privacy Act of 2018: artificial intelligence: training

DIGEST

This bill prohibits California Consumer Privacy Act (CCPA) covered-businesses that are the developers of artificial intelligence (AI) systems or tools from using the personal information of consumers under the age of 16 to train AI systems or services without first obtaining affirmative authorization, and even with such authorization the data must be de-identified and aggregated before it is used to train.

EXECUTIVE SUMMARY

The CCPA grants consumers certain rights with regard to their personal information, including enhanced notice, access, and disclosure; the right to deletion; the right to restrict the sale or sharing of information; and protection from discrimination for exercising these rights. (Civ. Code § 1798.100 et seq.) In the November 3, 2020 election, voters approved Proposition 24, which established the California Privacy Rights Act of 2020 (CPRA). The CPRA amends the CCPA, limits further amendment, and creates the California Privacy Protection Agency (PPA). The CCPA has special protections for children the business knows are under 16 years old, prohibiting the selling or sharing of their information without affirmative authorization, as provided.

In response to privacy concerns related to the training of AI systems, the bill amends the CCPA to prohibit using the personal information of the same population discussed above, unless the developer of the AI system or service gets affirmative authorization in the same manner as above. Even when such authorization is granted, the personal information shall be deidentified and aggregated before being used to train an AI system or service. This bill is sponsored by Common Sense Media. It is supported by Protection of the Educational Rights of Kids (PERK) Advocacy. The Committee received no timely opposition.

PROPOSED CHANGES TO THE LAW

Existing law:

1) Establishes the CCPA, which grants consumers certain rights with regard to their personal information, including enhanced notice, access, and disclosure; the right to deletion; the right to restrict the sale of information; and protection from discrimination for exercising these rights. It places attendant obligations on businesses to respect those rights. (Civ. Code § 1798.100 et seq.)

2) Provides a consumer the right, at any time, to direct a business that sells or shares personal information about the consumer to third parties not to sell or share the consumer's personal information. It requires such a business to provide notice to consumers, as specified, that this information may be sold or shared and that consumers have the right to opt out of that selling and sharing. (Civ. Code § 1798.120(a)-(b).)

3) Prohibits a business, notwithstanding the above, from selling or sharing the personal information of consumers if the business has actual knowledge that the consumer is less than 16 years of age, unless the consumer, in the case of consumers at least 13 years of age and less than 16 years of age, or the consumer's parent or guardian, in the case of consumers who are less than 13 years of age, has affirmatively authorized the sale or sharing of the consumer's personal information. A business that willfully disregards the consumer's age shall be deemed to have had actual knowledge of the consumer's age. (Civ. Code § 1798.120(c).)

4) Provides that the obligations imposed by the CCPA shall not restrict a business's ability to carry out certain conduct, including complying with federal, state, or local laws or to cooperate with law enforcement. This also includes cooperating with a government agency's request for emergency access to a consumer's personal information if a natural person is at risk or danger of death or serious physical injury where certain circumstances are met. (Civ. Code § 1798.145(a).)

5) Defines "personal information" as information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household. The CCPA provides a nonexclusive series of categories of information deemed to be personal information, including identifiers, biometric information, and geolocation data. (Civ. Code § 1798.140(v).) The CCPA defines and provides additional protections for sensitive personal information, as defined, that reveals specified personal information about consumers. (Civ. Code § 1798.140(ae).)

6) Establishes the CPRA, which amends the CCPA and creates the PPA, which is charged with implementing these privacy laws, promulgating regulations, and carrying out enforcement actions. (Civ. Code § 798.100 et seq.; Proposition 24 (2020).)

7) Permits amendment of the CPRA by a majority vote of each house of the Legislature and the signature of the Governor, provided such amendments are consistent with and further the purpose and intent of this act as set forth therein. (Proposition 24 § 25 (2020).)

This bill:

1) Provides that if the sale or sharing of a consumer's personal information requires affirmative authorization under subdivision (c) of Section 1798.120, a developer shall not use the personal information of that consumer to train an AI system or service unless the consumer or the consumer's parent or guardian has affirmatively authorized, in the same manner as above, that use of the consumer's personal information. If affirmative authorization is given, the personal information shall be deidentified and aggregated before being used to train an AI system or service.

2) States that the Legislature finds and declares that this act furthers the purposes and intent of the CPRA.

3) Defines the relevant terms, including:
    a) "Artificial intelligence" means an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.
    b) "Train" means exposing artificial intelligence to data in order to alter the relationship between inputs and outputs.
    c) "Developer" means a covered business that designs, codes, or produces an automated decision tool, or substantially modifies an artificial intelligence system or service for the intended purpose of making, or being a controlling factor in making, consequential decisions, whether for its own use or for use by a third party.

COMMENTS

1. California's landmark privacy protection law

As stated, the CCPA grants consumers certain rights with regard to their personal information, as defined. With passage of the CPRA in 2020, the CCPA got an overhaul. Consumers are afforded the right to receive notice from businesses at the point of

collection of personal information and the right to access that information at any time. The CCPA also grants a consumer the right to request that a business delete any personal information about the consumer the business has collected from the consumer. However, a business is not required to comply with such a request to delete if it is necessary for the business to maintain the consumer's personal information in order to carry out certain obligations or other conduct. (Civ. Code § 1798.105(d).)

The CCPA provides adult consumers the right, at any time, "to direct a business that sells personal information about the consumer to third parties not to sell the consumer's personal information. This right may be referred to as the right to opt-out." Changes made by the CPRA extend this to opting out of the "sharing" of the personal information as well. A business is thereafter prohibited from selling (or sharing) that information unless consent is subsequently provided. A business that sells or shares personal information to third parties is required to notify consumers that this information may be sold and that they have the right to opt out of such sales. (Civ. Code § 1798.120(b).) The CPRA added a new category of information, sensitive information, which includes data such as precise geolocation and genetic information. Consumers are additionally empowered to limit businesses' use of such information.

2. Protecting the integrity of consumer opt outs

Owing to recent advances in processing power and the rise of big data, AI's capacity and the scope of its applications have expanded rapidly, impacting every facet of our lives. Ultimately, AI systems are only as good as the data used to train them. AI system training is an iterative process whose success depends on the quality and depth of the input: "An AI model is a program that has been trained on a set of data to recognize certain patterns or make certain decisions without further human intervention. Artificial intelligence models apply different algorithms to relevant data inputs to achieve the tasks, or output, they've been programmed for."[1]

However, there is very little transparency in what data is used to train these systems and that lack of transparency hamstrings efforts to address and adequately identify many of the issues being raised by AI's rapid development.

One area of particular concern is the use of personal information to train AI systems and the privacy implications of such training. This bill tackles the use of children's personal information in the training of such models by placing restrictions on developers, defined as covered businesses that design, code, or produce an automated decision tool, or substantially modify an AI system or service for the intended purpose of making, or being a controlling factor in making, consequential decisions, whether for its own use or for use by a third party.

[1] IBM, *What is an AI model?*, https://www.ibm.com/topics/ai-model.

The CCPA currently provides heightened protections for children. If a business has actual knowledge that a consumer is less than 16 years of age, it is prohibited from selling or sharing the child's personal information, with one exception. If the business gets affirmative authorization from a child over 13, or the child's parent or guardian if they are younger than 13 years old, the business can sell or share the child's personal information. A business that willfully disregards the consumer's age shall be deemed to have had actual knowledge of the consumer's age.

This bill prohibits a developer from using the personal information of a consumer the developer knows is a child under 16 years of age to train an AI system or service, unless the consumer or their parent or guardian provides affirmative authorization for such use, in the same manner as provided in the CCPA. However, before using the information to train the model, the personal information must be deidentified and aggregated.

According to the author:

> As AI becomes ever more present in our lives, we must be especially cautious about AI designed to interact primarily with children and teenagers. Currently, businesses can use minors' personal data to train AI without any safeguards or the consent or even knowledge of the child or their parents. AB 2877 establishes rules around who can use children's personal data to better protect vulnerable young people from businesses that use their data to train AI systems.

3. Stakeholder positions

Common Sense Media, the sponsor of the bill, writes:

> While education and awareness about AI is valuable and necessary, the state must also ensure needed protections and guardrails are in place for kids, teens, and other users. When social media platforms launched 20 years ago, policymakers in California and the U.S., much like most people across the country, did not understand or foresee the consequences of this soon-to-be dominant technology. The results of that indifference are now clear, as research shows that, for young people in particular, the influence of social media platforms, the constant pull to be online 24/7, and the collection and sharing of vast amounts of data on young people has had, and continues to have, significant consequences for the mental and physical well-being of young users and broader aspects of all of our civic and social lives.
>
> Now, with the advent of AI and generative AI, we should all have learned our lesson. Again, there are clearly significant benefits but also great risks

> from AI products and services. Among the many risks, one of them is the potential to weaken or evade data privacy protections for minors. AB 2877 takes a simple but important step in ensuring that data collection through AI training models is aligned with the
> California Privacy Protection Act (CCPA).

The California Chamber of Commerce argues in opposition:

> Certainly, on its face, a law that prohibits developers from using children's data to train AI absent first obtaining affirmative authorization and then both deidentifying and aggregating their data, appears more privacy protective than the status quo. In practice, because individualized data can significantly enhance the performance, personalization, and accuracy of an AI model, it is entirely likely that such restrictions will undermine the ability of developers to train their AI to, among other things, be privacy protective. By that same token, it is also likely that such restrictions will significantly constrain the ability of developers to properly and adequately train their AI to avoid or mitigate certain outcomes, such as those that result in bias and discrimination. Of course, while Californians certainly have a constitutional right of privacy, they also are constitutionally protected against discrimination as well. Thus, by failing to balance competing interests and rights, Californians may not actually benefit from the additional restrictions placed by AB 2877 on the usage of minors PI.

In response to these concerns, the author has agreed to amendments that allow for use of such personal information if the training on such data is necessary to protect the consumer from imminent threats to their physical health or safety and the information is deidentified and aggregated prior to training. The amends also provide more clarity on the later training or fine-tuning that is done on models after the initial training.

Writing in support, Oakland Privacy asserts:

> The premise of Assembly Bill 2877 is that the infusion of personal data into an artificial intelligence system for the purposes of training the system *is* a sale or share of personal information under the California Privacy Rights Act.
>
> Given that information put into an artificial intelligence system to train it cannot be easily removed and remains in the system for the remainder of its practical life, and can be deployed for a variety of uses and applications, including some that may not have been anticipated in the training phase, this premise does not seem controversial. It does not upend the fundamental regulatory structure of the CCPA and CPRA and

the federal COPPA Act from whose language the guidelines in CPPA/CPRA relating to minor's personal information was drawn from.

It is clear to us that the training and development of artificial intelligence systems has not been COPPA and CPRA-compliant to date e.g. OpenAI and Google have confirmed using transcriptions of Youtube videos, a platform heavily used by minors and teens, to train their generative AI models.

4. Furthering the purpose and intent of the CPRA

Section 25 of the CPRA requires any amendments thereto to be "consistent with and further the purpose and intent of this act as set forth in Section 3." Section 3 declares that "it is the purpose and intent of the people of the State of California to further protect consumers' rights, including the constitutional right of privacy." It then lays out a series of guiding principles. These include various consumer rights such as:

- consumers should know who is collecting their personal information;
- consumers should have control over how their personal information is used; and
- consumers should benefit from businesses' use of their personal information.

Section 3 also includes a series of responsibilities that businesses should have. These include:

- businesses should specifically and clearly inform consumers about how they use personal information; and
- businesses should only collect consumers' personal information for specific, explicit, and legitimate disclosed purposes.

Section 3 also lays out various guiding principles about how the law should be implemented.

This bill strengthens protections for children and their personal information. Therefore, as it explicitly states, this bill "furthers the purposes and intent of the California Privacy Rights Act of 2020."

SUPPORT

Common Sense Media (sponsor)
Oakland Privacy
PERK Advocacy

OPPOSITION

California Chamber of Commerce
Technet

RELATED LEGISLATION

Pending Legislation:
SB 1223 (Becker, 2024) includes "neural data," as defined, within the definition of "sensitive personal information" for purposes of the CCPA. SB 1223 is currently in the Assembly Privacy and Consumer Protection Committee.

AB 1824 (Valencia, 2024) requires a business that assumes control of all or some part of a transferor business that includes the transfer of a consumer's personal information to comply with a consumer's direction to the transferor pursuant to the CCPA. AB 1824 is currently on the Senate Floor.

AB 1949 (Wicks, 2024) prohibits the collection, sharing, selling, using, or disclosing the personal information of minors without affirmative consent from either the minor or their parent or guardian, as provided. The bill provides for regulations to be promulgated by the PPA. AB 1949 is currently in this Committee.

AB 2013 (Irwin, 2024) requires developers of AI systems or services that are made available for Californians to use to post on their website documentation regarding the data used to train the system or service, including high-level summaries of the datasets used. AB 2013 is currently in this Committee.

AB 3048 (Lowenthal, 2024) requires that internet browsers include an opt-out preference signal allowing consumers interacting with businesses online to automatically exercise their right to opt-out of the selling and sharing of their personal information. AB 3048 is currently in this Committee.

Prior Legislation:

AB 947 (Gabriel, Ch. 551, Stats. 2023) included personal information that reveals a consumer's citizenship or immigration status in the definition of "sensitive personal information" for purposes of the CCPA.

AB 1194 (Wendy Carrillo, Ch. 567. Stats. 2023) provided stronger privacy protections pursuant to the CCPA where the consumer information contains information related to accessing, procuring, or searching for services regarding contraception, pregnancy care, and perinatal care, including abortion services.

AB 375 (Chau, Ch. 55, Stats. 2018) established the CCPA.

PRIOR VOTES:

Assembly Floor (Ayes 73, Noes 0)
Assembly Appropriations Committee (Ayes 15, Noes 0)
Assembly Privacy and Consumer Protection Committee (Ayes 11, Noes 0)

****