Date of Hearing: April 23, 2024

ASSEMBLY COMMITTEE ON PRIVACY AND CONSUMER PROTECTION
Rebecca Bauer-Kahan, Chair
AB 2877 (Bauer-Kahan) – As Amended April 18, 2024

AS PROPOSED TO BE AMENDED

SUBJECT: California Consumer Privacy Act of 2018: California Privacy Protection Agency

SYNOPSIS

Artificial intelligence (AI) uses algorithms – sets of rules – to transform inputs into outputs. In order to function, AI tools first need to be trained. During this period of machine learning, an AI tool is exposed to data and allowed to explore its structure. Machine learning tools develop their algorithms automatically. The process of exposing a machine learning tool to data is known as "training." The algorithm that a machine learning tool develops during training is known as its "model."

Just as humans cannot intentionally forget information they have learned, it is not currently possible to remove data from a trained AI tool. Unlike an Excel spreadsheet, which stores data in neat columns, AI tools store data in the connections between neurons in a deep neural network. Every connection is influenced by every piece of training data, and a large model like ChatGPT-4 is reported to have more than 1.7 trillion connections. It is not possible to specifically alter these connections in order to remove data without fundamentally changing the model; as a result, for data to be removed, the model must be retrained from scratch. ChatGPT-4 is estimated to have taken 4-7 months to train in the first place. Any legislative efforts seeking to grant Californians a "right to be forgotten" by AI tools should consider targeting training data before it has been incorporated into a model.

This bill intends to do that for children. Specifically, it prohibits California Consumer Privacy Act (CCPA) covered-businesses that are the developers of AI tools from collecting and using the personal information of consumers under the age of 16 to train AI tools without first obtaining express consent, and even with consent the data must be de-identified and aggregated.

According to the sponsors of the bill, Common Sense Media, "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." Under the CCPA currently, the restrictions for the personal information of consumers under the age of 16 pertain to the selling and sharing of that information, not its collection and use. For specific businesses that are developing AI tools, this bill ensures that young consumers' personal information is not collected and used to train this tools without first obtaining authorization.

SUMMARY: Prohibits a developer, as defined, from collecting and using the personal information of consumers under 16 years of age to train an artificial intelligence system. Specifically, **this bill**:

1) Defines the following terms in the California Consumer Privacy Act (CCPA):

a) "Artificial intelligence" to mean 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) "Developer" to mean a business that designs, codes, or produces an artificial intelligence system or service or that substantially modifies and existing artificial intelligence system or service by training on personal information.

c) "Training" or "trains" to mean exposing artificial intelligence to data in order to alter the relationship between inputs and outputs.

2) Prohibits a developer form using the personal information of a consumer under the age of 16 to train an artificial intelligence system or service unless a consumer's parent or guardian, if the consumer is under 13, or the consumer authorizes the collection and use of their personal information for this purpose.

3) Limits the use of young consumers' personal information, after the consumer allows its use, the data must be de-identified and aggregated.

EXISTING LAW:

1) Provides, pursuant to the California Constitution, that all people are by nature free and independent and have inalienable rights. Among these the fundamental right to privacy. (Cal. Const. art. I, § 1.)

2) States that the "right to privacy is a personal and fundamental right protected by Section 1 of Article I of the Constitution of California and by the United States Constitution and that all individuals have a right of privacy in information pertaining to them." Further states these findings of the Legislature:

a) The right to privacy is being threatened by the indiscriminate collection, maintenance, and dissemination of personal information and the lack of effective laws and legal remedies.

b) The increasing use of computers and other sophisticated information technology has greatly magnified the potential risk to individual privacy that can occur from the maintenance of personal information.

c) In order to protect the privacy of individuals, it is necessary that the maintenance and dissemination of personal information be subject to strict limits. (Civ. Code § 1798.1.)

3) Establishes the California Consumer Privacy Act. (Civ. Code §§ 1798.100-1798.199.100.)

4) Limits a business' collection, use, retention, and sharing of a consumer's personal information to that which is reasonably necessary and proportionate to achieve the purposes for which the personal information was collected or processed, or for another disclosed purpose that is compatible with the context in which the personal information was collected, and not further processed in a manner that is incompatible with those purposes. (Civ. Code § 1798.100(c).)

5) Prohibits a business from selling or sharing the personal information of a child that is 16 years of age or younger, if the business has actual knowledge of the child's age, unless the child, or the child's parent or guardian in the case of children less than 13 years old, has affirmatively authorized the sharing of selling of the personal information. (Civ. Code § 1798.120(c).)

6) Provides that consumers have the right, at any time, to direct a business that collects sensitive personal information about the consumer to restrict the use of that information to only that use which is necessary to perform the services or provide the goods reasonably expected by an average consumer who requests those goods or services. (Civ. Code § 1798.121(a).)

7) Establishes the California Privacy Protection Agency (Privacy Agency), vested with full administrative power, authority, and jurisdiction to implement and enforce the CCPA. The Privacy Agency is governed by a five-member board, with the chairperson and one member appointed by the Governor, and the three remaining members are appointed by the Attorney General, the Senate Rules Committee, and the Speaker of the Assembly. (Civ. Code § 1798.199.10.)

8) Defines the following terms under the CCPA:

a) "Business" means a for-profit entity that collects consumers' personal information, does business in California, and meets one or more of the following criteria:

i) It had gross annual revenue of over $25 million in the previous calendar year.

ii) It buys, receives, or sells the personal information of 100,000 or more California residents, households, or devices annually.

i) It derives 50% or more of its annual revenue from selling California residents' personal information. (Civ. Code § 1798.140(d).)

b) "Personal information" means 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. Personal information includes such information as:

i) Name, alias, postal address, unique personal identifier, online identifier, IP address, email address, account name, social security number, driver's license number, passport number, or other identifier.

ii) Commercial information, including records of personal property, products or services purchased, obtained, or considered, or other purchasing or consuming histories or tendencies.

iii) Biometric information.

iv) Internet activity information, including browsing history and search history.

v) Geolocation data.

vi) Audio, electronic, visual, thermal, olfactory, or similar information.

vii) Professional or employment-related information. (Civ. Code § 1798.140(v).)

c) "Sensitive personal information" means personal information that reveals a person's:

i) Social security, driver's license, state identification card, or passport number.

ii) Account log-in, financial account, debit card, or credit card number in combination with any required security or access code, password, or credentials.

iii) Precise geolocation.

iv) Racial or ethnic origin, citizenship or immigration status, religious or philosophical beliefs, or union membership.

v) Email, mail and text messages.

vi) Genetic data.

vii) Information collected and analyzed relating to health.

viii) Information concerning sex life or sexual orientation. (Civ. Code § 1798.140(ae).)

FISCAL EFFECT: As currently in print, this bill is keyed fiscal.

COMMENTS:

1) **Background.** Artificial intelligence (AI) uses algorithms – sets of rules – to transform inputs into outputs. Inputs and outputs can be anything a computer can process: numbers, text, audio, video, or movement. This is because AI is not fundamentally different from other computer functions. Its novelty lies in its application: unlike normal computer functions, AI is able to accomplish tasks that are normally performed by humans. Calculation and data storage are normal computer functions. Software that interprets emotions in human faces is AI.[1]

Machine learning. In order to function, AI tools first need to be trained. During this period of machine learning, an AI tool is exposed to data and allowed to explore its structure. Machine learning tools develop their algorithms automatically.[2] The process of exposing a machine learning tool to data is known as "training." The algorithm that a machine learning tool develops during training is known as its "model."

Models that are trained on small, specific datasets in order to make recommendations and predictions are sometimes referred to as "predictive AI." This differentiates them from "generative AI," which are trained on massive datasets in order to produce detailed text and

---

[1] Tate Ryan-Mosley, "AI isn't great at decoding human emotions. So why are regulators targeting the tech?," *MIT Technology Review,* August 14, 2023, https://www.technologyreview.com/2023/08/14/1077788/ai-decoding-human-emotions-target-for-regulators/.

[2] IBM, "What is machine learning?," www.ibm.com/topics/machine-learning.

.

images. When Netflix suggests a TV show to a viewer, the recommendation is produced by predictive AI that has been trained on the viewing habits of Netflix users.[3] When ChatGPT generates text in clear, concise paragraphs, it uses generative AI that has been trained on the entire written contents of the internet.[4] These systems' architecture and training data differ, but their core concept is the same.

Removing data from a trained model. Just as humans cannot intentionally forget information they have learned, it is not currently possible to remove data from a trained AI tool.[5] Unlike an Excel spreadsheet, which stores data in neat columns, AI tools store data in the connections between neurons in a deep neural network. Every connection is influenced by every piece of training data, and a large model like ChatGPT-4 is reported to have more than 1.7 trillion connections.[6] It is not possible to specifically alter these connections in order to remove data without fundamentally changing the model; as a result, for data to be removed, the model must be retrained from scratch. ChatGPT-4 is estimated to have taken 4-7 months to train in the first place.[7] Any legislative efforts seeking to grant Californians a "right to be forgotten" by AI tools should consider targeting training data before it has been incorporated into a model. This bill seeks to do that for children.

2) **Need for this bill.** In California, the CCPA secures increased privacy rights for consumers, including the right to know what personal information businesses collect and sell and the right to stop those sales to third parties. With respect to young users, the CCPA prohibits businesses from selling the personal information of a child that they know to be under the age of 16 unless they get affirmative authorization ("opt-in") for the sale of the child's personal information. Importantly, the current restrictions on children's information under the CCPA relate to the selling and sharing of personal information. It does not prohibit the collection and use of that information.

It is reasonable for consumers to expect that the personal information related to children that is being collected and used to train AI tools deserves the same strict protections as the selling and sharing of that same information. The parent or guardian, for the youngest consumer, or the consumer themselves for teenagers, should be able to prohibit the use of their personal information and authorize its use for specific purposes with specific restrictions.

The purpose of this bill is to prohibit CCPA covered-businesses that are the developers of AI tools from collecting and using children's personal information to train AI tools without first obtaining express consent and even with consent, the data must be de-identified and aggregated.

---

[3] Netflix, "How Netflix's Recommendations System Works," help.netflix.com/en/node/100639.

[4] OpenAI, "How ChatGPT and Our Language Models Are Developed," help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed.

[5] Stephen Pastis, "A.I.'s un-learning problem: Researchers say it's virtually impossible to make an A.I. model 'forget' the things it learns from private user data," *Yahoo! Finance*, August 30, 2023, finance.yahoo.com/news/un-learning-problem-researchers-virtually-164342971.html.

[6] Reed Albergotti, "Microsoft pushes the boundaries of small AI models with big breakthrough," *SEMAFOR*, November 1, 2023, www.semafor.com/article/11/01/2023/microsoft-pushes-the-boundaries-of-small-ai-models.

[7] Stephen McAleese, "Retrospective on 'GPT-4 Predictions' After the Release of GPT-4," *LESSWRONG*, March 17, 2023, https://www.lesswrong.com/posts/iQx2eeHKLwgBYdWPZ/retrospective-on-gpt-4-predictions-after-the-release-of-gpt.

3) **Author's statement.** According to the author:

> As AI becomes ever more present in our lives, we must be especially cautious about AI that is designed to primarily interact with children. Businesses can use children's personal data to train AI without any safeguards and without the consent or even knowledge of the child or their parents. With AB 2877 I am committed to establishing 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.

4) **The California Consumer Privacy Act and the California Privacy Rights Act (CPRA).** In 2018, the Legislature enacted the CCPA (AB 375 (Chau, Chap. 55, Stats. 2018)), which gives consumers certain rights regarding their personal information, such as the right to: (1) know what personal information about them is collected and sold; (2) request the categories and specific pieces of personal information the business collects about them; and (3) opt out of the sale of their personal information, or opt in, in the case of minors under 16 years of age.

Subsequently, in 2020, California voters passed Proposition 24, the California Privacy Rights Act (CPRA), which established additional privacy rights for Californians. With the passage of the CCPA and the CPRA, California now has the most comprehensive laws in the country when it comes to protecting consumers' rights to privacy.

In addition, Proposition 24 created the California Privacy Protection Agency (Privacy Agency) in California, vested with full administrative power, authority, and jurisdiction to implement and enforce the CCPA and the CPRA. The Agency's responsibilities include updating existing regulations, and adopting new regulations.

To protect Californians from any future legislative efforts to weaken statutory protections in the CPRA, Proposition 24 provided that the CPRA's contents may be amended by a majority vote of the Legislature only if the amendments are consistent with and further the purpose and intent of the CPRA, which is to further protect consumers' rights, including the constitutional right of privacy.[8]

The provisions contained in this bill clearly further the goals of the CPRA by increasing restrictions on the collection and use of young people's personal information.

5) **Proposed Committee amendments.** The suggested Committee amendments are intended to clarify the bill and create consistency within the CCPA.

Amendment #1: The definitions included in this bill will be moved to the general definitions section of the CCPA, rather than being confined to the new section proposed by this bill.

Amendment #2: The author of the bill, who chairs this Committee, has proposed, in another bill, a standard definition of AI that is intended to be used throughout the state's statutes. The definition is updated in this bill to reflect that definition:

> "Artificial intelligence" means an ~~engineered or machine-based system that, for explicit or implicit objectives, infers from the input it receives how to generate outputs that can~~

---

[8] Ballot Pamphlet. Primary Elec. (Nov. 3, 2020) text of Prop. 24, p. 74

~~influence physical or virtual environments. 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.*

Amendment # 2: The bill as it is currently in print includes in the definition of "developer" parties that are not covered by the CCPA. In order to address this inconsistencies, the amendments do the following:

"Developer" means a ~~person, partnership, state or local government agency, or corporation~~ *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.

Amendment #3: Adds the definition of "training" and "trains" to the CCPA:

"Training" or "trains" to mean exposing artificial intelligence to data in order to alter the relationship between inputs and outputs.

Amendment #4: Consistent with the current age demarcations in the CCPA, the following definition will be deleted, and instead, the first sentence in (b) will be updated to reflect the CCPA covered entities. In addition (b) removes the reference to "sensitive" in order to broaden the scope of the bill to include all personal information:

~~(3) "Minor" means a natural person under 18 years of age.~~

(b) A developer shall not use ~~a minor's sensitive~~ the personal information of a consumer whose personal information is restricted under 1798.120(c) to train an artificial intelligence system or service.

Amendment #5: In addition to the above amendments, the author understands that there are specific instances when it is important to have access, with permission, to the personal information of children for the training of beneficial AI. The following will be added to (b) in amendment #4:

. . , unless the consumer or consumer's parent or guardian, as described in 1798.120(c), authorizes its collection and use. If authorization is given, the data must be de-identified and aggregated before being used to train an artificial intelligence system or service.

5) **Related legislation.** Over the last 5 years numerous bills have attempted to modify the CCPA and many have been successful in furthering its goals. This year alone, this Committee has heard five bills, including this one, that propose modifications to the CCPA. Specifically:

AB 1824 (Valencia, 2024) requires, under the CCPA, that businesses that are acquiring the personal data of consumers through the acquisition of another business, honor the previous decisions of consumers who have not given permission for the business to sell or share their personal information. That bill is currently pending in the Appropriations Committee.

AB 1949 (Wicks, 2024) amends the CCPA to prohibit a business from collecting the personal information of a consumer under 18 years of age unless the consumer, or the consumer's parent or guardian if under 13, affirmatively authorizes the collection. That bill is currently pending in the Appropriations Committee.

AB 3048 (Lowenthal, 2024) proposes prohibiting a business from developing or maintaining an internet browser through which a consumer interacts with a business that does not include a setting that enables the consumer to send an opt-out preference signal to that business. That bill is currently pending in the Appropriations Committee.

AB 3204 (Bauer-Kahan, 2024) requires data digesters to register with the Privacy Agency. In addition, defines a "data digester" as an organization or enterprise, including, but not limited to, a proprietorship, partnership, firm, business trust, joint venture, syndicate, corporation, association, or nonprofit that designs, codes, or produces an artificial intelligence system or service, or that substantially modifies an existing artificial intelligence system or service, by training the system or service on the personal data of 1,000 or more individuals or households. This bill is currently pending before this Committee.

ARGUMENTS IN SUPPORT:

Common Sense Media, the sponsors of this bill, write in support:

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).

REGISTERED SUPPORT / OPPOSITION:

Support

Common Sense Media
Perk Advocacy

Opposition

None on file.

Analysis Prepared by: Julie Salley / P. & C.P. / (916) 319-2200