

From the Privacy & Cybersecurity Group of Weil, Gotshal & Manges LLP

November 10, 2025

AI Transparency Data Disclosures: Preparing for California AB 2013

Beginning **January 1, 2026**, companies that design, code, produce, or significantly update certain generative AI systems must publicly disclose on their websites information concerning the datasets used to train those generative AI systems pursuant to [Cal. AB 2013](#), or the California Generative Artificial Intelligence Training Data Transparency Act.

For businesses working to determine whether Cal. AB 2013 applies to them and prepare initial drafts of the required disclosures, our questionnaire and template below offer a good place to start. Please contact olivia.greer@weil.com, eliza.cotter@weil.com, or max.scott@weil.com with any questions on completing this questionnaire, assistance with populating the template disclosure, to obtain sample disclosures prepared by our team, or otherwise for assistance complying with the requirements of Cal. AB 2013.

Cal. AB 2013 Questionnaire

Part A. Is Our Company Subject to the Requirements of Cal. AB 2013?

1. Have we designed, coded or produced a generative AI system¹ that we (or a third party) have made publicly available?
2. Have we retrained, fine-tuned or made any other significant update to a generative AI system that changes its functionality or performance that we (or a third party) have made publicly available?

If the answer to both questions in Part A is *no*, your company is not subject to the requirements of Cal. AB 2013.

If the answer to either question in Part A is *yes*, please move to Part B to determine whether an exception is applicable.

Part B. Is Our Generative AI System Subject to Any Exceptions under Cal. AB 2013?

1. Are any of the following statements true about this generative AI system?
 - a. The system's sole purpose is to help ensure security and integrity;²
 - b. The system's sole purpose is the operation of aircraft in the national airspace;
 - c. The system was developed for national security, military, or defense purposes and is made available only to a U.S. federal government entity.

If the answer to *any* of the subparts in Part B is *yes*, the generative AI system is not within the scope of Cal. AB 2013.

If the answer to *each* of the subparts in Part B is *no*, your company is subject to the requirements of Cal. AB 2013 and you should move to Part C to determine the information required for your company's website disclosure.

¹ "Generative artificial intelligence" refers to AI that can generate synthetic content, such as text, images, video, and audio.

² For purposes of this questionnaire, "security and integrity" means the ability to detect security incidents, resist malicious, deceptive, fraudulent, or illegal actions and to help prosecute those responsible for such actions, and to ensure physical safety.

Part C. What Information Should We Disclose Regarding Our Generative AI System?

1. For each dataset (i.e., each single, pre-packaged collection of data) used to test, validate or fine-tune the generative AI system subject to Cal. AB 2013,³ please:
 - a. Identify the source or owner of the dataset and indicate whether the dataset was purchased or licensed;
 - b. Provide the time period when the data in the dataset were collected, indicate whether collection is ongoing, and identify when the dataset was first used in the development of the generative AI system;
 - c. Provide a general range for the number of data points included in the dataset (e.g., for text, in tokens; for images, in numbers of images; for video, in hours of video content), including an estimated figure if the dataset is a dynamic dataset.
 - d. Confirm whether the dataset contains:
 - i. Data protected by copyright, trademark or patent law (or, if applicable, indicate whether the data is entirely in the public domain, and therefore not subject to copyright, trademark or patent law);
 - ii. Personal information;⁴
 - iii. Aggregate consumer information;⁵
 - iv. Synthetic data.⁶
 - e. Confirm whether we cleaned, processed or otherwise modified the dataset:
 - i. If so, please describe the intended purpose of the cleaning, processing or modification of the dataset.
 - f. Describe:
 - i. If the dataset includes labels, the types of labels used;
 - ii. If the dataset does not include labels, the general characteristics of the data;⁷
 - iii. With respect to both (i) and (ii), as applicable, how the dataset will contribute to the purpose of the generative AI system.⁸

³ For pre-existing AI models that we fine-tuned, distilled or otherwise modified, these questions should be answered with respect to the training content we used and our *own* modifications of the generative AI system, not the training content or process used by the underlying AI model provider.

⁴ "Personal information" means information that identifies or is reasonably capable of being associated or linked with a particular consumer or household.

⁵ "Aggregate consumer information" means information that relates to a group or category of individuals, from which individual identities have been removed such that the information is not reasonably linkable to any individual or household, including via a device.

⁶ "Synthetic data" refers to data generated when seed data are used to create artificial data that have some of the statistical characteristics of the seed data.

⁷ General characteristics may include the format (e.g., image, audio, video, text) and sample values of the underlying data points. Sample values will depend on the format of the data: (i) for image, examples may include photography, visual art works, infographics, social media images, logos, or brands; (ii) for audio, examples may include musical compositions and recordings, audiobooks, radio shows and podcasts, or private audio communications; (iii) for video, examples may include music videos, films, TV programs, performances, video games, video clips, journalistic videos, or social media videos; (iv) for text, examples may include fiction and non-fiction text, scientific text, press publications, legal and official documents, social media comments, or source code.

⁸ This purpose explanation can be relatively high-level (e.g., "because the dataset is comprised of images of trees, it will help our AI system, which is intended to identify objects in nature, achieve its intended purpose", or "because the dataset is comprised of guitar sounds, it will help our AI system, which is intended to create music based on specific genres").

Cal. AB 2013 Template Disclosure⁹

System Information¹⁰

- **System Name:**
- **Developer:** [*Company Name*]
- **Substantial Modification:**¹¹ [Yes] or [No]
 - **[Model Dependencies and Developer, as applicable:** [indicate underlying model/version and developer]]
- **Release Date:** [*MM/DD/YYYY*]
- **Version:** [*e.g., v1.0, v2.5*]

⁹ Cal. AB 2013 requires companies that design, code, produce, or significantly update certain generative AI systems or services (the “Systems”) to publicly disclose on their websites information concerning the datasets used to train those generative AI systems (the “Datasets”). Systems that are exempted from this disclosure requirement are Systems (i) with the sole purpose of helping ensure security and integrity; (ii) with the sole purpose of operating aircraft in the national airspace; or (iii) that were developed for national security, military, or defense purposes and are made available only to a U.S. federal government entity.

¹⁰ For pre-existing Systems (i.e., third-party Systems) that the disclosing company has made a “substantial modification” (defined below) to, these questions should be answered with respect to the training content the *disclosing company* used and such company’s modifications of the pre-existing System, not the Datasets used by the developer of the pre-existing System to train the pre-existing System.

¹¹ “Substantial Modification” refers to a new version, new release, or other update to a System that materially changes its functionality or performance, including the results of retraining or fine tuning.

High-Level Summary of Datasets

*[Dataset 1] [Group of Datasets 1]*¹²

Key Information	[Template Disclosure]			
Source or Owner	<i>[Name of source of the Dataset] or [Name of owner of the Dataset]</i>			
Purchased or Licensed	<i>[This Dataset was [purchased or licensed] or [N/A]]</i>			
Time Period of Collection of Data in Dataset	<i>[Month, YYYY – Month, YYYY] or [Data in this Dataset was originally collected beginning in [YYYY]; data collection for this Dataset is currently ongoing]</i> ¹³			
Date of First Use in Development of System	<i>[Month, YYYY]</i>			
Number of Data Points	<i>[Approximately [value and unit of measurement of data points in the Dataset - e.g., 5,000,000 tokens, 10,000 images, 100 recordings]]</i> ¹⁴			
Contains Data Protected by Copyright, Trademark, or Patent Law	<i>[Yes] or [No]</i> ¹⁵			

¹² The template disclosures do not necessarily need to be provided for *each* specific Dataset used to train the relevant System (although, if the System was trained on a limited number of Datasets, this may be easiest). For Systems that were trained on a high number of Datasets, grouping disclosures by common characteristics of the Datasets may be the preferred approach (e.g., Datasets consisting of English speech recordings and Datasets consisting of Spanish speech recordings could be grouped together in a set of disclosures regarding speech recordings).

¹³ Use the second disclosure option if the collection of data in the Dataset is continuous/ongoing.

¹⁴ For a dynamic Dataset, a general range of data points is acceptable (e.g., less than 10,000 data points, between 10,000-1,000,000 data points, over 1,000,000 data points).

¹⁵ If the Dataset contains data protected by copyright, trademark, or patent law, the disclosing company may wish to provide additional context regarding such data, including how the relevant intellectual property rights have been secured by the disclosing company. The determination as to whether additional context or clarification to this disclosure is appropriate will depend on the facts and circumstances relating to the Dataset, the System, and the disclosing company.

Key Information	[Template Disclosure]			
Dataset is Entirely in the Public Domain	[Yes] or [No]			
Contains Personal Information	[Yes] or [No] ¹⁶			
Contains Aggregate Consumer Information	[Yes] or [No]			
Uses Synthetic Data	[Yes] or [Yes, the purpose of the <i>[use or continuous use]</i> of the synthetic data in this Dataset is to <i>[explain the functional need of the use of the synthetic data in relation to the purpose of the System - e.g., train safety filters of the System]</i> ¹⁷ or [No]			
Dataset Has Been Cleaned, Processed or Otherwise Modified	[Yes, this Dataset was <i>[high-level description of cleaning, processing, and/or other modification activity - e.g., filtered for profanity]</i> in order to <i>[high-level description of purpose of modification activity in relation to System - e.g., maintain platform standards]</i>			
Description of Types of Data Points	<i>[List the types of labels used in the Dataset or, if labels are not used in the Dataset, include a general description of the format of the Dataset and sample values- e.g., images, including images of trees and lakes]</i>			
Purpose in Relation to System	[This Dataset furthers the intended purpose of the System by <i>[general description of how the Dataset helps the System achieve its purpose - e.g., improving the System's reasoning]</i>			

¹⁶ If the Dataset contains personal information, the disclosing company may wish to provide additional context regarding the type of personal information contained in the Dataset, including whether such personal information is publicly available. The determination as to whether additional context or clarification to this disclosure is appropriate will depend on the facts and circumstances relating to the Dataset, the System, and the disclosing company.

¹⁷ The statute provides that developers *may* (but are not required to) include a description of the functional need or desired purpose of the synthetic data in relation to the intended purpose of the system or service.

Privacy & Cybersecurity Alert is published by the Privacy & Cybersecurity practice group of Weil, Gotshal & Manges LLP, 767 Fifth Avenue, New York, NY 10153, +1 212 310 8000, www.weil.com.

If you have questions concerning the contents of this issue, or would like more information about Weil's Privacy & Cybersecurity practice group, please speak to your regular contact at Weil or to authors:

Authors

Liza Cotter (BO)	View Bio	eliza.cotter@weil.com	+1 617 772 8365
Olivia Greer (NY)	View Bio	olivia.greer@weil.com	+1 212 310 8815
Max Scott (SF)	View Bio	max.scott@weil.com	+1 415 470 5130
Taya Bokert (NY)	View Bio	taya.bokert@weil.com	+1 212 310 8321
Brendan McNerney (NY)	View Bio	brendan.mcnerney@weil.com	+1 212 310 8815

© 2025 Weil, Gotshal & Manges LLP. All rights reserved. Quotation with attribution is permitted. This publication provides general information and should not be used or taken as legal advice for specific situations that depend on the evaluation of precise factual circumstances. The views expressed in these articles reflect those of the authors and not necessarily the views of Weil, Gotshal & Manges LLP. If you would like to add a colleague to our mailing list, please [click here](#). If you need to change or remove your name from our mailing list, send an email to weil.alerts@weil.com