

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

August 20, 2025

Weil Draft Internal Questionnaire for Cal. AB 2013 Disclosures

California has passed [Cal. AB 2013](#), or the California Generative Artificial Intelligence Training Data Transparency Act, a groundbreaking new bill set to take effect January 1, 2026. Cal. AB 2013 requires companies that design, code, produce, or significantly update certain generative AI systems to publicly disclose on their websites information concerning the datasets used to train those generative AI systems. Our questionnaire below is a good place to start to help determine (1) whether your company is subject to the requirements of Cal. AB 2013 and (2) if so, what information to include in your company's website disclosure. Please contact olivia.greer@weil.com, eliza.cotter@weil.com, or max.scott@weil.com with any questions on completing this questionnaire or otherwise complying with the requirements of Cal. AB 2013.

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 in 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 those actions, and 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 data⁶.
 - 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 and
 - 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, other) 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, private audio communication; (iii) for video, examples may include music videos, films, TV programs, performances, video games, video clips, journalistic videos, 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, and 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.")

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 8496

© 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