The Computer & Internet Lawyer

Volume 43 ▲ Number 6 ▲ June 2026

Ronald L. Johnston, Arnold & Porter, Editor-in-Chief

California's GenAI Data Training Compliance Law AB 2013: Challenges and Practical Next Steps for Protecting IP

By Marguerite McConihe and Sam Cohen

California's new Generative Artificial Intelligence law, AB 2013,[1] ushers in a new era of transparency for developers of generative AI (GenAI) systems. Signed into law on September 28, 2024 and effective on January 1, 2026, this legislation requires organizations who utilize GenAI systems to publish clear, high-level details about the datasets used in the training or development of the systems.

Whether an organization is simply using GenAI models or is actively involved in the development of such models, the law's broad requirements create onerous compliance requirements that businesses need to implement quickly, and present significant and complex challenges to protecting proprietary intellectual property assets.

Developing effective strategies and procedures to comply with this law (and the many bills pending in state legislatures with similar requirements or regarding other aspects of AI regulation) may seem daunting. Collaborating with experienced intellectual property counsel will facilitate the transition to compliance with this new law.

The authors, attorneys with Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., may be contacted at mmcconihe@mintz.com and slcohen@mintz.com, respectively.

WHO IS COVERED?

The law casts a wide net. At a high level, AB 2013 requires developers to publicly post information about their training data. A "developer" under the law is not limited to GenAI model developers, but also organizations that "substantially modify" them.

The law applies to any system, whether free or paid, that is "available to Californians" and was either released or significantly modified on or after January 1, 2022. Under the law, a "substantial modification" to an AI system includes any new versions or releases or any other material changes to an AI system's functionality or performance. Any organization that designs, produces, fine-tunes, retrains, or otherwise substantially alters AI systems or service must comply. Each of these actions falls under the broad definition of the term "developer."

WHAT MUST BE DISCLOSED?

Developers subject to AB 2013 must publish a "high-level summary" detailing the datasets used in the development and training of their systems. These summaries are required to include information such as:

- Sources / owners of datasets and how each dataset furthers the intended purpose.
- Approximate volume (ranges / estimates for dynamic data).
- Types of data points (labels used or general characteristics for unlabeled data).
- IP status (presence of copyright, trademark, patent, or entirely publicdomain data), and whether datasets were purchased / licensed.
- Whether datasets include personal information or aggregate consumer information (as defined in the CCPA / CPRA).
- Whether the datasets contain synthetic information.
- Any cleaning, processing, or modifications performed, and their intended purpose.

There may be additional granularity required, but given the law's nascence it remains unclear the level of detail needed to comply with each category of information. This raises the need for a substantial compliance overhaul and, at minimum, the need for record-keeping throughout the development and evaluation periods of GenAI systems.

COMPLIANCE CHALLENGES AND STRATEGIC TIPS

AB 2013 introduces several significant compliance challenges for developers of GenAI systems, largely due to the statute's broad and sometimes ambiguous language. One of the primary concerns involves the impact the disclosure requirements may have on the value of IP assets, primarily trade secrets, as well as raising significant concerns for a company's IP strategy, particularly with regard to copyrights and patents.

The law requires developers to reveal the details of the training data used in their GenAI systems. This information is often highly sensitive and contains valuable proprietary and trade secret material owned by AI developers.

This challenge is addressed in a new lawsuit filed by xAI on December 29, 2025. In its complaint, xAI argues that AB 2013 forces organizations to disclose proprietary training datasets. According to the document, this allegedly violates the Constitution's Taking Clause, compels speech in breach of the First Amendment, and violates due process due to vagueness and ambiguity. Although the lawsuit is in early stages, this is one case to track for developments in AB 2013 compliance.

Another key challenge in complying with AB 2013 lies in its requirement that businesses explicitly state whether their datasets include copyrighted or patented material, even if only as training inputs. This raises additional IP-related risks, among them the creation of litigation risks by signaling to third-party rights holders that their IP is part of the training data. It also undermines the business's potential position in copyright or patent infringement cases by creating an impossible standard whereby the businesses can be argued to have always been placed "on notice." These requirements could effectively compel disclosing the business's competitive edge.

Developers must also navigate the complexities of identifying and accounting for personal or consumer information, particularly under California regulations such as the CCPA / CPRA, when datasets originate from third-party or open-source repositories, which may lack transparency into the nature and origin of the data. It is as yet unclear whether the law creates an independent requirement to verify the data source.

Industry critics have argued that the law's extensive scope, covering a wide array of AI systems, imposes burdens that exceed those seen in other jurisdictions, potentially stifling innovation and increasing regulatory overhead for organizations already subject to multiple compliance regimes. However, as was the case with California's data protection laws, GenAI disclosure legislation will likely soon start popping up around the country.

Given these challenges, organizations must consider a strategic approach to compliance. This includes reassessing data governance policies, implementing robust compliance frameworks, and ensuring active involvement of legal teams throughout the development life cycle.

Balancing the need for transparency with the risks to reputation and regulatory exposure is crucial, particularly when handling sensitive or proprietary information. To ensure long-term compliance, organizations should also seek alignment with evolving U.S. and international standards, recognizing that regulatory

expectations for AI systems are likely to become more stringent and harmonized over time.

EXEMPTIONS

AB 2013 includes a few exemptions, but they are drawn narrowly. Systems used solely for security or integrity purposes, for aircraft operation in national airspace, or developed for national security / military / defense are not covered by this statute.

PRACTICAL NEXT STEPS

Organizations must be proactive when planning their compliance strategy. The following steps can provide a simple roadmap; however, consulting with trusted legal counsel is always the best strategy.

- Inventory all GenAI systems (public-facing and internal) released or modified since January 2022. Identify clear ownership and licensing agreements for each business IP asset, flagging any dataset that lacks documentation.
- Develop website disclosure templates addressing each requirement.
- Maintain documentation of data sourcing and processing to support disclosures, creating specific public-facing versions that mitigate the risk of trade secret disclosure.
- Confer with legal counsel to find new ways to keep proprietary and trade secret information confidential, including potentially renegotiating any license or confidentiality agreement terms.
- Build a legally defensible IP position *before* making any disclosures, creating, with the help of legal counsel, a litigation response and enforcement risk plan.
- Monitor legal developments, including in the *xAI* case, that argue the law forces disclosure of proprietary training datasets and other potential regulatory shifts.

WHY IT MATTERS

- Failure to comply could lead to enforcement by the California AG or other state entities.
- Enhanced transparency aims to address consumer trust, bias, copyright, and privacy concerns.
- This law is a potential template for other states or even federal AI regulation. Addressing a strategy for compliance now will save organizations time and headaches later when enforcements become more common.

Note

1. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB2013.