

# From AI Doomers to E/Accs: How SB 1047 and the 38 AI Laws in California Are Shaping Future AI Law

California's Safe and Secure Innovation for Frontier Artificial Intelligence Models Act is one of the first significant regulations of artificial intelligence in the United States that, if signed, would place liability on the developers of AI models.

By The Honorable Jerry McNerney, Elizabeth Vella Moeller, Jeewon K. Serrato, Shruti Bhutani Arora, Amaris Trozzo

## TAKEAWAYS

- Ⓢ Before a company can begin to initially train a covered model, the bill would require AI developers to publicly post disclosures about how the company will test the likelihood of the model to cause critical harm and the conditions under which the model will be fully shutdown, as well as a filing of the same with the California Attorney General (AG).
- Ⓢ Violations would be enforceable by the California AG with the civil penalty up to 10% of the cost of the quantity of computing power used to train the model and 30% for any subsequent violation.
- Ⓢ SB 1047 would also establish a Government Operations Agency, a consortium that would be required to develop a framework for the creation of a public cloud computing cluster.

---

09.26.24

The California legislature sent 38 AI bills to the Governor's office as the 2024 legislation session came to a close, eight of which have already been signed, regulating everything from deepfake nudes and AI-generated celebrity clones to election tampering. Governor Newsom has until September 30 to sign the

rest, including California Senate Bill 1047, known as the Safe and Secure Innovation for Frontier Artificial Intelligence Models Act. SB 1047 is one of the first significant regulations of artificial intelligence in the United States that, if signed, would place liability on the developers of AI models.

While the bill is designed to help prevent catastrophic harms, the law, as is currently drafted, would apply to all large-scale models, regardless of the potential for harm. The law would apply to models that cost at least \$100 million to train. Before a company can begin to initially train a covered model, the bill would require AI developers to publicly post disclosures about how the company will test the likelihood of the model to cause critical harm and the conditions under which the model will be fully shutdown, as well as a filing of the same with the California Attorney General (AG). Violations would be enforceable by the California AG with the civil penalty up to 10% of the cost of the quantity of computing power used to train the model and 30% for any subsequent violation. SB 1047 would also establish a Government Operations Agency, a consortium that would be required to develop a framework for the creation of a public cloud computing cluster known as “CalCompute” for use by developers and researchers statewide.

What happens in California doesn’t just stay in California; it often sets the stage for nationwide standards. This is why so many eyes are now on SB 1047.

The law would apply to models that cost at least \$100 million to train.

SB 1047 may not impact all AI systems that are currently in service but will certainly have a significant impact on how large AI developers behave. There will be significant testing, governance, and reporting hurdles for companies to jump through before any large-scale AI models can initiate training. Because much of the reporting is required to be done publicly, this may have the impact of chilling certain AI developers. In his public remarks on September 17, Newsom discussed how he is considering not only the “outsized impact” SB 1047 will have in the AI industry but also potentially the “chilling effect.”

Whether you are an “AI doomer,” believing that AI presents an existential threat to humanity, or you are an “effective accelerationist” or “e/accs” and believe that AI can bring a utopian future, it is clear AI regulations are here and now, perhaps trying to address not only the demonstrable risks but also to anticipate the hypothetical risks. SB 1047 highlights the ongoing debate around where the responsibility sits—whether AI laws should regulate the models and the developers or the applications and uses of the AI technology.

With a significant concentration of AI talent and a majority of AI companies headquartered in California, this bill is likely to influence not just state-level, but national and global AI policy.

### **Key Provisions**

SB 1047 regulates “covered model” and “covered model derivative.”

“Covered model” is defined as models that are:

- Trained using a quantity of computing power greater than  $10^{26}$  integer or floating-point operations, the cost of which exceeds \$100 million; or
- Created by fine-tuning a covered model using a quantity of computing power equal to or greater than three times  $10^{25}$  integer or floating-point operations, the cost of which, as reasonably assessed by the developer, exceeds \$10 million if calculated using the average market price of cloud compute at the start of fine-tuning. On and after January 1, 2027, the Government Operations Agency may adopt a different definition for models that are fine-tuning a covered model.

“Covered model derivative” means an unmodified copy of a covered model, a copy of a covered model that has been subjected to post-training modifications unrelated to fine-tuning, as well as a covered model that has been fine-tuned and a copy of a covered model that has been combined with other software.

SB 1047 requires AI developers of covered models to institute several measures to ensure the safe, responsible use of their systems, including:

### **Pre-Training Requirements**

- Developers must implement comprehensive safety and security measures to protect the training process of the model, including protections from misuse, unauthorized access and unsafe post-training alterations. These security measures, which must be documented, can be tiered to the risk presented by the type of model.
- Developers will need to retain an unredacted copy of the safety and security protocol for as long as the covered model is made available for commercial, public, or foreseeably public use plus five years, including records and dates of any updates or revisions, and also conduct an annual review of the safety and security protocol to account for any changes to the capabilities of the covered model and industry best practices and, if necessary, make modifications to the policy.
- Written protocols will also need to describe a testing procedure which would evaluate whether a covered model poses an unreasonable risk of causing or enabling a critical harm, specify how the developer will protect against “unreasonable risk of causing or materially enabling a critical harm,” address the possibility that a covered model or covered model derivative can be used to make post-training modifications or create another covered model in a manner that may cause or materially enable a critical harm, and the conditions under which a developer would enact a full shutdown.
- Developers must also institute capabilities to promptly shut down the model if needed.

### **Prohibition on Harmful Use**

- The bill explicitly prohibits the deployment of AI models that could pose an unreasonable risk of causing or enabling critical harm.
- Developers are required to ensure that users cannot access the “hazardous capabilities” of the model or cause a “critical harm.” Any potential harms must be able to be traced back to the user.

- The key to understanding these requirements on developers to avoid harm is the definition of a “critical harm,” which is met where AI-enables the creation of a chemical, biological, radiological or nuclear weapon (that results in mass casualties); AI is used to carry out a cyberattack targeting critical infrastructure resulting in \$500 million in damages or mass casualties; or the use of AI in mass casualties or \$500 million in damages when the AI (a) acts with limited human oversight and (b) results in grave in physical damage (and if carried out by a human could be considered a crime).

### **Audit Requirements**

- Starting January 1, 2026, all large-scale AI developers, not just those working on models that could cause critical harm, would be required to undergo independent third-party audits annually to ensure compliance with SB 1047’s provisions. These audit reports must be retained for the duration of the model’s commercial or public use, plus an additional five years. A redacted copy of the auditor’s report must be publicly published (again, presumably on the company’s website) and transmitted to the AG.

### **Certification and Reporting**

- Developers will need to retain an unredacted copy of the safety and security protocol for as long as the covered model is made available for commercial, public, or foreseeably public use plus five years, including records and dates of any updates or revisions, and also conduct an annual review of the safety and security protocol to account for any changes to the capabilities of the covered model and industry best practices and, if necessary, make modifications to the policy.
- Developers must conspicuously publish a copy of the redacted safety and security protocol and transmit a copy of the redacted safety and security protocol to the California AG. The AG will be able to request access to a copy of the unredacted safety and security protocol.
- A developer of a covered model will need to annually submit to the AG a statement of compliance signed by the chief technology officer, or a more senior corporate officer, for as long as the covered model or any covered model derivatives controlled by the developer remain in commercial or public use or remain available for commercial or public use. Among others, this annual statement by the officer must include a description of the process used by the signing officer to verify compliance with the requirements of this law, including a description of the materials reviewed by the signing officer, a description of testing or other evaluation performed to support the statement and the contact information of any third parties relied upon to validate compliance.
- An AI safety incident, defined as an incident that “demonstrably increases the risk of a critical harm occurring,” must be reported no later than 72 hours following learning that the incident has occurred or learning facts “sufficient to establish a reasonable belief” that an incident has occurred. AI safety incidents would include: (1) a covered model or covered model derivative autonomously engaging in behavior other than at the request of a user; (2) theft, misappropriation, malicious use, inadvertent release, unauthorized access, or escape of the model weights of a covered model or covered model derivative; (3) the critical failure of technical or administrative controls, including controls limiting the ability to modify a covered model or covered model derivative; and (4) unauthorized use of a covered model or covered model derivative to cause or materially enable critical harm.

## Computing Cluster Requirements

Operators of computing clusters (defined as “a set of machines transitively connected by data center networking of over 100 gigabits per second that has a theoretical maximum computing capacity of at least  $10^{20}$  integer or floating-point operations per second and can be used for training artificial intelligence”) would need to:

- obtain the prospective customer’s basic identifying information and business purpose for utilizing the computing cluster,
- assess whether the prospective customer intends to utilize the computing cluster to train a covered model,
- if a customer repeatedly utilizes computer resources that would be sufficient to train a covered model, validate the information initially collected and conduct the assessment prior to each utilization,
- retain a customer’s Internet Protocol addresses used for access or administration and the date and time of each access or administrative action,
- maintain for seven years and provide to the AG, upon request, appropriate records of actions taken under this section, including policies and procedures put into effect, and
- implement the capability to promptly enact a full shutdown of any resources being used to train or operate models under the customer’s control.

Additionally, SB 1047 would require:

- a person that operates a computing cluster to consider industry best practices and applicable guidance from the S. Artificial Intelligence Safety Institute, National Institute of Standards and Technology and other reputable standard-setting organizations,
- that a person who operates a computing cluster may impose reasonable requirements on customers to prevent the collection or retention of personal information that the person that operates a computing cluster would not otherwise collect or retain, including a requirement that a corporate customer submit corporate contact information rather than information that would identify a specific individual, and
- the obligations on computing clusters could affect startups that use computing clusters for their AI modeling.

## Penalties for Non-Compliance

- Developers found in violation may face the suspension of their AI model’s commercial or public use until they can demonstrate full compliance with the bill’s safety and security protocols.
- The California AG can impose fines amounting to up to 10% of the cost of the quantity of computing power used to train a covered AI model, with penalties increasing to 30% for subsequent violations.
- The bill also allows for the recovery of monetary damages, punitive damages, attorney’s fees and injunctive relief, making the financial implications of non-compliance significant.

- A provision within a contract or agreement that seeks to waive, preclude or burden the enforcement of a liability arising from a violation of this chapter, or to shift that liability to any person or entity in exchange for their use or access of, or right to use or access, a developer's products or services, including by means of a contract of adhesion, would be void as a matter of public policy.
- The bill also provides that a court shall disregard corporate formalities and impose joint and several liability on affiliated entities.

### **Employee Notice and Whistle Blower Protections**

- SB 1047 protects employees who report to the AG any reasonable belief that the company they work for is out of compliance with the requirements of SB 1047. The employer is prohibited from retaliating against any such employee on these grounds.
- Developers must provide an internal process by which employees can anonymously bring issues of noncompliance to the attention of the employer.
- A developer must provide a clear notice to all employees working on covered models and covered model derivatives of their rights and responsibilities under the law, including the right of employees of contractors and subcontractors to use the employer's anonymous whistle blower hotline. Employers must not only post and display this notice within all workplaces but also provide a written copy and receive acknowledgement of receipt from all employees annually.

### **California Passes Eight AI Laws**

While many of the AI bills, including SB 1047 still remain pending signature on his desk, Newsom has already signed eight AI bills into law, tackling the areas of deepfake nudes, political uses of AI, celebrity AI clones and AI watermarking:

- SB 926 makes it a criminal act to blackmail someone with AI-generated nude images;
- SB 981 requires social media platforms to establish reporting channels for deepfake nudes;
- AB 2602 requires studios to enter into a written contract with an actor before creating an AI-generated digital replica of a performer's voice or likeness, and the performer must be professionally represented in negotiating the contract;
- AB 1836 prohibits studios from using digital replicas of deceased performers without first obtaining the consent of those performers' estates;
- AB 2355 would require disclosures about AI-generated political advertisements. Our article about the Federal Communications Commission's (FCC) proposed rules with similar disclosure requirements can be found [here](#);
- [AB 2655](#) requires large online platforms to remove or label AI deepfakes related to elections, as well as create channels to report such content. Candidates and elected officials can seek injunctive relief if a large online platform is not complying with the act;
- [AB 2839](#) takes aim at social media users who post, or repost, AI deepfakes that could deceive voters about upcoming elections; and

- SB 942 requires widely used generative AI systems to add watermarks to AI-generated content saying they are AI generated.

### What's Next

If SB 1047 is signed, it would set the precedent for how societies balance safety with innovation. We will publish another update after September 30 to review which California AI bills were signed into law and how each of them may interact with one another. Among the bills awaiting the Governor's signature is AB 2013, legislation that would require developers to post disclosures on their websites about datasets used to train a generative AI system or service. An article with an overview of AB 2013's requirements can be found [here](#).

While 2024 was a very active legislative year, this is only the beginning. Many more regulations are expected to be pushed in 2025. Businesses should evaluate what AI tools and services they currently deploy. Whether you are a developer of models or purchaser of AI systems, companies should develop a comprehensive compliance strategy, including, but not limited to, conducting risk assessments that identify requirements depending on jurisdiction and different usage applications.

Companies that take a proactive approach to how they harness the power of AI technologies will emerge as leaders in the field, driving responsible innovation and setting industry standards. Given how fast evolving the regulatory landscape is around the world, companies should also actively monitor and perhaps even participate in the legislative and regulatory process to influence the future of AI governance.

For more information or to discuss how SB 1047 and other emerging AI regulations and legislation may impact your business, please contact us.

*Pillsbury's multidisciplinary team of AI thought leaders and legal and strategic advisors is an industry leader in strategic promotion of responsible and beneficial AI. Pillsbury is closely monitoring AI-related legislative and regulatory efforts. Our AI team helps startups, global corporations and government agencies navigate the landscape impacted by emerging developments in AI. For insights on these rapidly evolving topics, please visit our [Artificial Intelligence practice page](#).*

These and any accompanying materials are not legal advice, are not a complete summary of the subject matter, and are subject to the terms of use found at: <https://www.pillsburylaw.com/en/terms-of-use.html>. We recommend that you obtain separate legal advice.