

Insights: Alerts

Proposed California AI Law SB 1047 – An Overview for Developers

September 12, 2024

Written by **Stephen M. Anstey** and **Michael J. Breslin**

California AI [Senate Bill 1047](#), titled the Safe and Secure Innovation for Frontier Artificial Intelligence Models Act (“SB 1047” or the “bill”), has the potential to significantly impact AI development and use. The bill, which passed the California Assembly on August 28, 2024, and the California Senate on August 29, 2024, has been sent to Governor Newsom and awaits his action. If signed into law, SB 1047 would represent the most comprehensive AI specific state legislation enacted to date.

SB 1047 is a complex and far-reaching proposed statute. The below is an overview of those provisions and aspects of the bill that concern developers and likely to be relevant to most companies.

1. Definitions

SB 1047 provides key definitions for several AI specific terms. In many instances these proposed definitions differ from those of other significant AI focused legislation and regulation, including the [EU AI Act \(which entered into force on August 1, 2024\)](#).

“Covered model” and “fine-tuning”

Broadly, a “covered model” under SB 1047 includes, “(i) An artificial intelligence model trained using a quantity of computing power greater than 10^{26} integer or floating-point operations, the cost of which exceeds one hundred million dollars (\$100,000,000) when calculated using the average market prices of cloud compute at the start of training as reasonably assessed by the developer[; and] (ii) An artificial intelligence model 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 ten million dollars (\$10,000,000) if calculated using the average market price of cloud compute at the start of fine tuning.”

Fine-tuning is defined as “adjusting the model weights of a trained covered model or covered model derivative by exposing it to additional data.”

Takeaways

SB 1047's inclusion of fine-tuning is unique and expands the definition beyond initial model development. The inclusion of associated model costs is also a unique aspect of the provided definition. That said, training models, whether initial training or subsequent fine-tuning, is an extremely complex task, and in practice it may prove

difficult to ascertain exact costs. Finally, there are no current AI models that meet the designated floating-point operations requirement, so this definition is prospective.

“Critical harm”

“Critical harm” enjoys an expansive definition under SB 1047, and includes, “any of the following harms caused or materially enabled by a covered model or covered model derivative.”

- “The creation or use of a chemical, biological, radiological, or nuclear weapon in a manner that results in mass casualties.
- Mass casualties or at least five hundred million dollars (\$500,000,000) of damage resulting from cyberattacks on critical infrastructure by a model conducting, or providing precise instructions for conducting, a cyberattack or series of cyberattacks on critical infrastructure.
- Mass casualties or at least five hundred million dollars (\$500,000,000) of damage resulting from an artificial intelligence model engaging in conduct that does both of the following:
 - Acts with limited human oversight, intervention, or supervision.
 - Results in death, great bodily injury, property damage, or property loss, and would, if committed by a human, constitute a crime specified in the Penal Code that requires intent, recklessness, or gross negligence, or the solicitation or aiding and abetting of such a crime.
- Other grave harms to public safety and security that are of comparable severity to the harms described [above].”

The bill also includes examples of what “critical harm” does not include, which are not covered here.

Takeaways

While all the articulated harms are relevant, one could most easily perceive instances where a model could cause at least \$500 million in damage from property loss, though the requirement that the model act with “limited human oversight, intervention, or supervision” makes the standard more difficult to meet. What has the potential to be most concerning to effected companies though in the catchall provision, which includes “other grave harms to public safety and security that are of comparable severity to the harms described.”

“Developer”

Under the bill, a developer means “a person that performs the initial training of a covered model either by training a model using a sufficient quantity of computing power and cost, or by fine-tuning an existing covered model or covered model derivative using a quantity of computing power and cost greater than the amount specified [i.e., \$100 million and \$10 million respectively].”

Takeaways

The developer definition is rather straightforward but important to address given the broad meaning provided for covered models and the significant number of companies that are likely to qualify as developers under SB 1047.

2. Before training a covered model

Developers are required to meet various requirements before initially training a covered model. Developers must:

- Prevent unauthorized access to and access of the model,
- Implement the capability to fully shut down the model,
- Draft and retain a safety and security protocol, testing procedures, and information on how the developer is meeting SB 1047's requirements,
- Conduct an annual review of the necessary protocols,
- Take reasonable care to implement measures to prevent the models from posing unreasonable risk, and
- Conspicuously publish a copy of the redacted safety and security protocol and transmit a copy of the redacted safety and security protocol to the Attorney General.

Takeaways

The requirements developers must meet before they train a covered model are considerable. Most notably, developers may bristle at having to implement a full shutdown system, perform an annual review, and draft and provide the Attorney General with their unredacted safety and security protocol before commencing initial training.

3. Before using a covered model or making it publicly available

SB 1047 requires that developers before using a covered model or covered model derivative (broadly defined as an unmodified copy of a covered model, or a covered model that has been subjected to post-training modifications unrelated to fine-tuning) in specific circumstances or making a model available for use, must:

- Assess whether the covered model is reasonably capable of causing or materially enabling a critical harm,
- Record necessary related data for third-party testing,
- Take reasonable care to implement appropriate safeguards to prevent the covered model and covered model derivatives from causing or materially enabling a critical harm, and
- Take reasonable care to ensure that the models' actions, as well as critical harms resulting from their actions, can be accurately and reliably attributed to them.

Takeaways

This provision primarily concerns harm and steps an entity will take to limit harm. One aspect that could be particularly impactful is requiring a developer to retain data regarding a model's actions and allowing such actions to be attributed to the model.

4. Use of covered models

The bill states that “[a] developer shall not use a covered model or covered model derivative for a purpose not exclusively related to the training or reasonable evaluation of the covered model or compliance with state or federal law or make a covered model or a covered model derivative available for commercial or public, or foreseeably public, use, if there is an unreasonable risk that the covered model or covered model derivative will cause or materially enable a critical harm.”

Takeaways

This provision is rather self-explanatory and again addresses potential risk. One interesting aspect of this provision, and others throughout the bill, is that requirements are affected if the developer follows another

relevant “state or federal law.” The U.S. has not yet passed a comprehensive federal AI law but may do so in the future.

5. Third-party audits, audit reports, and statements of compliance

Beginning January 1, 2026, developers are required to retain a third-party auditor to perform an independent audit to address compliance with the bill. The auditor will draft a related audit report and the developer will publish a redacted copy of the audit report and provide the Attorney General with access to an unredacted copy upon request. Separate from and in addition to the audit report, a developer will annually submit to the Attorney General a statement of compliance with the bill.

Takeaways

Requiring that an annual third-party audit be performed, and the results published, is a significant requirement for developers (both practically and legally). Given the novelty of emerging AI technologies, including generative AI, and the fact that the bill necessarily covers only the most advanced models, it will be difficult to determine what auditors are qualified to provide these audits, what standards they will meet, the form the reports will take, and the broader impact of these annual reports.

6. Reporting Safety Incidents

SB 1047 requires that a “developer of a covered model shall report each artificial intelligence safety incident affecting the covered model, or any covered model derivatives controlled by the developer, to the Attorney General within 72 hours of the developer learning of the artificial intelligence safety incident or within 72 hours of the developer learning facts sufficient to establish a reasonable belief that an artificial intelligence safety incident has occurred.” The bill defines an “artificial intelligence safety incident” as “an incident that demonstrably increases the risk of a critical harm occurring by means of any of the following: (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.”

Takeaways

In some ways the definition of what constitutes a safety incident is narrow enough to limit the compliance requirements for incident reporting. For example, models should necessarily be designed not to autonomously engage in problematic behavior and critical failures should be limited. That said, misappropriation or malicious use of model weights or unauthorized use of a model to cause or materially enable critical harm are, in a world full of bad actors, more likely to occur. With this in mind, requiring developers to report a safety incident within 72 hours of the developer learning of the incident or facts sufficient to establish a reasonable belief in the incident sets a burdensome obligation for developers to meet.

7. Civil actions, fines, and whistleblowers

The Attorney General may bring a civil action for violation of the bill. As to penalty amounts, the bill stipulates for “causes death or bodily harm to another human, harm to property, theft or misappropriation of property, or that

constitutes an imminent risk or threat to public safety[...] a civil penalty in an amount not exceeding 10 percent of the cost of the quantity of computing power used to train the covered model [...] for a first violation and in an amount not exceeding 30 percent of that value for any subsequent violation.” The Attorney General may also seek injunctive or declaratory relief, monetary damages, attorney's fees and costs, and “any other relief that the court deems appropriate.” The bill also provides protections for related whistleblowers, and holds that a developer may not “prevent an employee from disclosing information to the Attorney General or the Labor Commissioner” relating to failure to comply with the bill or concerning a model which “poses an unreasonable risk of causing or materially enabling critical harm, even if the employer is not out of compliance with any law.” Employer's must inform employees of these whistleblower provisions and develop processes to allow employees to anonymously disclose relevant information.

Takeaways

Given the current and growing utilization of models across sectors and the stated instances under which civil action may be brought, this provision grants the Attorney General broad authority and discretion to bring suit. The costs associated with training a model can be considerable, especially in initial training, so penalties sought may be significant. Additionally, extensive whistleblower protections further compound the gravity of a civil suit brought by the Attorney General concerning SB 1047.

Conclusion

If signed into law, SB 1047 would represent the most comprehensive and farthest-reaching AI state legislation enacted to date. The bill's broad scope and stringent requirements, including the comprehensive definition of covered models, risk determinations, safety protocols, reporting standards, audit requirements, and support for Attorney General civil actions all would have significant impact on AI developers. Entities that may be impacted by the bill should monitor its progress closely and prepare to take all necessary steps to comply if it is enacted into law.

Kilpatrick Connect – AI Legal Consulting

There is no development as consequential or with more legally significant implications for your business as the recent advancements in AI. Kilpatrick Connect is a legally focused AI consulting and advisory offering built upon Kilpatrick's AI, legal, and industry expertise and delivered through a confidential attorney-client relationship. We understand the transformative capabilities of AI and its profound impact on your business, and Kilpatrick Connect provides a safe, secure, and economical hub to address AI-related questions, issue resolution, and strategy development.

For more information on Kilpatrick Connect, please visit our website, [Kilpatrick Connect – AI Legal Consulting](#).

Related People



Stephen M. Anstey

t 202.824.1427

sanstey@ktslaw.com



Michael J. Breslin

t 404.685.6752

mbreslin@ktslaw.com