View on our website.

GIBSON DUNN

Artificial Intelligence Update September 25, 2024

Regulating the Future: Eight Key Takeaways from California's SB 1047, Pending with Governor Newsom

The bill purports to regulate only the most powerful AI models, trained using large computing capacity, but its requirements are likely to have a broader impact, including on open source models.

On August 28, 2024, the California State Assembly passed proposed bill SB 1047, the Safe and Secure Innovation for Frontier Artificial Intelligence Models Act, through which California seeks to regulate foundational AI models and impose obligations on companies that develop, fine-tune or provide compute resources to train such models.

SB 1047 currently sits with Governor Newsom. As of September 24, it is unclear whether the Governor will sign the bill or veto it; on September 17, Newsom signaled some discomfort with the bill, but stated that he remained undecided even as he signed several other AI-related bills into law.[1] Gov. Newsom has until the end of September to sign or veto the bill; if he does not veto or return the bill to the legislature, SB 1047 will become law and take effect on January 1, 2026, even if he does not sign it.

Controversial since its introduction, SB 1047 represents a major shift in how U.S. states have sought to regulate AI to date, and the novel approach–including its requirements for developers to implement a "kill switch" and subject themselves to third-party compliance audits, and its

applicability to startups and open source AI developers–has caused many major players in the technology sector to oppose the bill or work to weaken its provisions.

Below are 8 key takeaways that highlight the most important aspects of SB 1047 and the ways it may shape the AI landscape if it becomes law.

1. **Expansive definitions of "covered models" and "covered model derivatives" are likely to capture many frontier AI models and subsequent modifications.** SB 1047 broadly applies to "covered models," which are AI models that either:
    - Cost over $100 million to develop and are trained using computing power "greater than 10^26 integer or floating-point operations" (FLOPs); or
    - Are based on covered models and fine-tuned at a cost of over $10 million and using computing power of three times 10^25 integer or FLOPs.[2]

The frontier models that are publicly available are just below the covered AI model threshold, but the next generation of models will most likely hit that regulation mark.

Certain of SB 1047's requirements also apply to "covered model derivatives," which include copies of covered models (whether or not they have been modified).

2. **SB 1047's requirements apply only to companies that develop or provide compute power to train covered models or covered model derivatives, <u>not</u> to companies that merely use covered models.** The law's principal requirements apply to "developers" that initially train a covered model or that fine-tune a covered model or covered model derivative, all based on the applicable cost and compute requirements. Additional requirements apply to operators of computing clusters when one of their customers "utilizes compute resources that would be sufficient to train a covered model[.]"
3. **Before training a covered model, developers are required to implement technical and organization controls designed to prevent covered models from causing "critical harms."** These critical harms include creating or using certain weapons of mass destruction to cause mass casualties; causing mass casualties or at least $500 million in damages by conducting cyberattacks on critical infrastructure or acting with only limited human oversight and causing death, bodily injury, or property damage in a manner that would be a crime if committed by a human; and other comparable harms.
    - **Kill switch or "shutdown capabilities."** Developers are required to implement a means through which to "promptly enact a full shutdown" of all covered models and covered model derivatives in their control, such that all model operations, including further training, are stopped. In determining whether to enact a full shutdown, developers are required to consider whether it may cause any potential disruptions to critical infrastructure.
    - **Cybersecurity protections.** Developers are required to implement protections "appropriate in light of the risks" to prevent unauthorized access, misuse, or "unsafe post-training modifications" of the covered model and all covered model derivatives in their control.
    - **Safety protocols.** Developers are required to develop a written document safety and security protocol (SSP) and to designate a senior individual to implement the SSP in a manner that complies with the developer's obligation to exercise

reasonable care to mitigate the risk of "foreseeable" downstream misuse of covered models, including by reviewing the SSP for sufficiency on an annual basis. Developers are required to retain an unredacted version of their SSP for the life of the covered model to which it applies plus 5 years, publish a redacted version of the SSP, and to provide an unredacted version to the Attorney General upon request. The SSP is required to:

- Specify the means through which the developer will comply with its duty to exercise reasonable care as set out above and describe in detail how the developer will comply with SB 1047;
- Describe how the SSP may be modified;
- Describes when the developer would implement a full shutdown;
- Set out testing procedures to determine whether the covered model and its derivatives pose an unreasonable risk of causing or enabling a critical harm or whether the covered model and its derivatives may be modified in a manner that poses such a risk; and
- States the developer's compliance obligations in sufficient detail to allow the developer or a third party to determine whether the SSP has been followed.

4. **Developers are subject to rigorous testing, assessment, reporting, and audit obligations.**
    - **Testing and Assessment.** Before using a covered model or making it publicly available, a developer is required to assess, including through testing as set out in the SSP, whether there is a possibility that the model could cause critical harm and to record and retain test results from these assessments such that third-parties are capable of duplicating these tests.
    - **Audits and Reports.** Beginning in 2026, developers are required to retain a third-party auditor to perform an independent, annual audit of their compliance with SB 1047. Developers are required to publish redacted copies of their audit reports and to provide unredacted copies to the Attorney General on request. The bill further requires developers to submit annual compliance statements to the Attorney General and to report safety incidents within 72 hours of discovery.
5. **Compute providers are required to implement policies and procedures for customers that use compute sufficient to train a covered model.** These procedures are required to include the ability to enact a full shutdown of compute used to train covered models, collecting and verifying identifying information for any customer that uses compute sufficient to train a covered model and assessing whether the customer intends to use the compute resources to train a covered model. Such information is required to be retained for 7 years and shall be provided to the Attorney General on request.
6. **Developers are prohibited from preventing employees from reporting noncompliance internally, to the Attorney General, or to the Labor Commissioner and may not retaliate against employees who do so.** These whistleblower protections include requirements that developers inform any employee or contractor working on covered models of their rights and to retain any complaints or reports made by employees or contractors for 7 years. Developers also are required to develop processes through which employees or contractors may make internal reports on an anonymous basis.

7. **Enforcement is exclusively by the Attorney General and does not include a private right of action.** The Attorney General may bring a civil action for violations of the bill that cause death or bodily harm; damage, theft, or misappropriation of property; or imminent public safety risks. The Attorney General may seek civil penalties, monetary damages (including punitive damages), injunctive or declaratory relief. Civil penalties for certain violations are capped at 10% of the cost of computing power used to train the covered model.

8. **Certain provisions of SB 1047 may be vulnerable to legal challenge based on constitutional principles.** While many of the bill's provisions will likely pass constitutional muster, including those requiring developers to take technical steps in relation to their covered models, SB 1047 remains subject to legal challenge based on its extraterritorial reach and its assessment requirements.

    - **No nexus to California.** SB 1047 does not have any textual nexus requiring that developers be located in California nor any requirements that covered models be developed, trained, or offered in California for the provisions to apply, standing in opposition to the general presumption that state laws do not apply outside of that state's borders.

    - **Assessments may violate the First Amendment.** The bill's assessment provisions may be subject to legal challenge that they are unconstitutional government mandates for developers to create speech, in violation of the First Amendment. The likelihood of such challenges may be increased by the Ninth Circuit's latest holdings that similar assessment provisions in California's Age-Appropriate Design Code Act and AB 587 (relating to social media platforms) are facially unconstitutional on First Amendment grounds.[3]

[1] *See* Jeremy B. White, *Gavin Newsom signals concerns about major AI safety bill*, Politico (Sept. 17, 2024), https://subscriber.politicopro.com/article/2024/09/gavin-newsom-signals-concerns-about-major-ai-safety-bill-00179727 (setting out Newsom's concerns that the bill may create a "chilling effect" and make it harder for California to maintain its status as the home of tech innovation).

[2] The proposed computing threshold mirrors the Biden administration's Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.

[3] *NetChoice v. Bonta*, No. 23-2969 (9th Cir. Aug. 16, 2024); *X Corp. v. Bonta*, No. 24-271 (9th Cir. Sept. 4, 2024).

The following Gibson Dunn lawyers assisted in preparing this update: Christopher Rosina, Frances Waldmann, Emily Maxim Lamm, Cassandra Gaedt-Sheckter, Vivek Mohan, and Eric Vandevelde.

Gibson, Dunn & Crutcher's lawyers are available to assist in addressing any questions you may have regarding these issues. Please contact the Gibson Dunn lawyer with whom you usually work, the authors, or any leader or member of the firm's Artificial Intelligence practice group:

Christopher Rosina – New York (+1 212.351.3855, crosina@gibsondunn.com)
Frances A. Waldmann – Los Angeles (+1 213.229.7914,fwaldmann@gibsondunn.com)
Keith Enright – Palo Alto (+1 650.849.5386, kenright@gibsondunn.com)
Cassandra L. Gaedt-Sheckter – Palo Alto (+1 650.849.5203, cgaedt-sheckter@gibsondunn.com)
Vivek Mohan – Palo Alto (+1 650.849.5345, vmohan@gibsondunn.com)
Robert Spano – London/Paris (+33 1 56 43 13 00, rspano@gibsondunn.com)
Eric D. Vandevelde – Los Angeles (+1 213.229.7186, evandevelde@gibsondunn.com)

Attorney Advertising: These materials were prepared for general informational purposes only based on information available at the time of publication and are not intended as, do not constitute, and should not be relied upon as, legal advice or a legal opinion on any specific facts or circumstances. Gibson Dunn (and its affiliates, attorneys, and employees) shall not have any liability in connection with any use of these materials. The sharing of these materials does not establish an attorney-client relationship with the recipient and should not be relied upon as an alternative for advice from qualified counsel. Please note that facts and circumstances may vary, and prior results do not guarantee a similar outcome.

If you would prefer NOT to receive future emailings such as this from the firm, please reply to this email with "Unsubscribe" in the subject line.

If you would prefer to be removed from ALL of our email lists, please reply to this email with "Unsubscribe All" in the subject line. Thank you.

© 2024 Gibson, Dunn & Crutcher LLP. All rights reserved. For contact and other information, please visit us at gibsondunn.com