SENATE THIRD READING
SB 53 (Wiener)
As Amended September 02, 2025
Majority vote

SUMMARY

Requires large developers of the most advanced, costly artificial intelligence (AI) systems to implement certain protocols and publically disclose the protocols they use to mitigate the risk of catastrophic harms. Requires operators of computing clusters to obtain specified information relating to customers. Provides for whistleblower protections and enforcement by the Attorney General (AG). Provides, upon appropriation, for the creation of a framework to create a public cloud computing cluster.

Major Provisions
1) Requires large developers who train AI models on a specified amount of compute and with revenues above $500 million to write, implement, comply with, and clearly and conspicuously publish on its internet website a safety and security protocol which details how the developer will address catastrophic harms and the protocols in place to address the materialization of such harms.

2) Requires a large developer, before or at the time of making a new foundation model available, to publish on their internet website a transparency report for the model that describes the risk assessments or risk mitigation assessments used by the developer during the development of the model, including whether the developer used third-parties in the assessments and whether any denoted risk thresholds were attained.

3) Requires the Attorney General to establish a mechanism for large developers to report critical safety incidents that have materialized as a result of the use of their models.

4) Permits the Attorney General to redefine the definition of a large developer beginning in 2027, which reflects the technological developments, scientific literature, national and international standards, as well as stakeholder engagement.

5) Upon appropriation, establishes in the Government Operations Agency a consortium required to develop a framework for the creation of a public cloud computing cluster to be known as "CalCompute" that advances the development and deployment of AI, as prescribed.

6) Prohibits a developer from making, adopting, enforcing, or entering into any rule, regulation, policy, or contract that prevents an employee from disclosing, or retaliating against an employee for disclosing, information to the Attorney General, a federal authority, a person with authority over the employee, or another employee who has the authority to investigate the issue, if the employee has reasonable cause to believe that the information discloses either of the following:

a) The developer's activities pose a critical risk.

b) The developer has made false or misleading statements about its management of critical risk.

COMMENTS

*Foundation models, SB1047, and what this bill does. "*Frontier" or "foundation" models are the largest and most powerful AI systems in development. These models are highly generalizable and have the potential to drive breakthroughs in science and medicine, streamline complex processes, and strengthen the economy. However, they also pose serious catastrophic risks due to their immense capabilities. A frontier model could help cure disease or design the next pandemic. It could automate bureaucratic functions or become autonomous and disrupt critical infrastructure.

Critics of AI safety argue that the evidence base is still too limited to justify regulation, and that prematurely imposing safeguards could stifle innovation. Responding to this dilemma, the International AI Safety Report, led by one of the "godfathers of AI," Yoshua Bengio, has warned:

> On the one hand, pre-emptive risk mitigation measures based on limited evidence might turn out to be ineffective or unnecessary. On the other hand, waiting for stronger evidence of impending risk could leave society unprepared or even make mitigation impossible, for instance if sudden leaps in AI capabilities, and their associated risks, occur.

In the 2024 legislative session, SB 1047 (Wiener) sought to address concerns surrounding frontier models by establishing a regulatory framework intended to prevent catastrophic harms. The bill would have required frontier model developers to create and implement comprehensive safety and security protocols before initiating training, to implement shutdown capabilities, and to perform risk assessments on models and implement reasonable safeguards, subject to third-party auditing, before releasing the models. The bill also would have prohibited releasing or using models that pose an unreasonable risk of catastrophic harms. Finally, SB 1047 would have created a new state agency – the Board of Frontier Models – to oversee the development of these models.

In vetoing the bill, Governor Gavin Newsom acknowledged the risks but emphasized the importance of addressing the evidence dilemma:

> Let me be clear—I agree with the author—we cannot afford to wait for a major catastrophe to occur before taking action to protect the public. California will not abandon its responsibility. Safety protocols must be adopted. Proactive guardrails should be implemented, and severe consequences for bad actors must be clear and enforceable. I do not agree, however, that to keep the public safe, we must settle for a solution that is not informed by an empirical trajectory analysis of AI systems and capabilities. Ultimately, any framework for effectively regulating AI needs to keep pace with the technology itself.

Following the veto, the Governor convened the Joint California Policy Working Group on AI Frontier Models, which used empirical research, historical case studies, and modeling to craft a policy framework for regulating frontier models. The Working Group published its final report in June 2025.

This bill seeks to implement the recommendations of the Working Group Report. Narrower than its predecessor, SB 53 takes a light-touch approach that focuses on transparency as the means of ensuring safety and accountability for the largest developers of the most powerful models. Large developers who harness an extraordinarily high amount of compute power must create,

implement, and publish both a safety and security protocol and a transparency report for each released model. The bill does not prescribe any particular standards for these plans: it simply requires developers to explain whether and how they assess, mitigate, and manage catastrophic risks – those that would result in more than 50 deaths or $1 billion in damage.

The bill also establishes a critical incident reporting mechanism, administered by the Attorney General (AG), to ensure that severe or high-risk events are tracked and addressed in a timely manner. Incidents must be reported within 15 days. The AG is further authorized to update the definition of a "large developer" through rulemaking to ensure that the bill remains responsive to technological advancements. Additionally, the bill establishes a consortium within the Government Operations Agency to create a public computing cluster, known as CalCompute, to support AI research and safety testing. The bill also provides whistleblower protections for employees and contractors of large developers who report risks or noncompliance. The AG is authorized to enforce the bill by seeking an unspecified civil penalty.

Furthermore, the bill incorporates additional recommendations from the Working Group Report. First, the bill requires developers to include capability thresholds in both their safety and security protocols and transparency reports. Developers must disclose when their models exceed these thresholds and must document any mitigation measures taken in response. Second, to avoid over-burdening smaller startups, the amendments add to the definition of "large developer" an annual revenue threshold of $500 million, in addition to the current compute threshold. Third, the bill requires developers to disclose whether they retain the ability to shut down a model under their control in the event of a critical incident. This is a disclosure requirement only; unlike SB 1047, this bill does not mandate that a developer maintain the technical ability to shut down a model. Fourth, if a critical incident presents an imminent threat, the bill requires the developer to report the incident within 24 hours to the appropriate law enforcement authority. Fifth, the bill directs the AG to publish an annual, anonymized, and aggregated summary of all critical incident and whistleblower reports. Ultimately, this would establish a baseline transparency regime for the largest developers of foundation models.

For a full analysis please see the policy committee analysis.

According to the Author
In 2024, as part of his veto of Senate Bill 1047 (Wiener), Governor Newsom's Joint California Working Group on AI Frontier Models was established – a group of top experts tasked with charting a course forward on AI policy for the developers of the most advanced AI systems. Their final report, released in June 2025, emphasized the growing evidence for risk of severe harm, such as "AI-enabled hacking or biological attacks, and loss of control" and argued "California has a unique opportunity to continue supporting developments in frontier AI while addressing substantial risks that could have far-reaching consequences for the state and beyond."

Drawing recommendations from Governor Newsom's working group report, Senate Bill 53 requires covered developers to write, implement, and publish their safety and security protocol in redacted form to protect intellectual property. It would also require covered developers to report certain, carefully defined critical safety incidents to the Attorney General and would allow members of the public to report incidents.

SB 53 only applies to AI companies that have trained a model with 10^{26} floating point operations (FLOPs), a measure of computational power. These companies are spending hundreds of millions of dollars to train the most advanced AI models. As recommended by the Report, SB 53

also authorizes the Attorney General to adjust the scoping of the bill in the future to keep up with technological developments, but only focuses on well-resourced AI companies at the frontier of AI development.

Senate Bill 53 strengthens whistleblower protections for employees of frontier artificial intelligence laboratory companies whose activities pose a catastrophic risk. SB 53 also establishes a consortium to help create CalCompute: a public AI research cluster that will provide startups and researchers with access to the resources needed to develop large-scale AI systems.

In doing this, and building on Governor Newson's working group report, SB 53 allows California to continue to maintain its leadership in the AI development ecosystem and to demonstrate that safety does not stifle success.

Arguments in Support
Secure AI Project, co-sponsors of the bill, alongside a coalition of technology equity advocacy groups write in support:

> The California Report on Frontier AI Policy, while it does not endorse any specific legislation, forms the foundation for SB 53. Established by Governor Newsom in 2024 and led by Dr. Fei-Fei Li, Dr. Jennifer Tour Chayes and Mariano-Florentino Cuéllar, the report is anchored on the notion of "trust but verify" and calls for more transparency into the safety practices of AI companies, adverse event reporting requirements, and whistleblower protections. SB 53 implements these principles.
>
> Large AI developers are developing increasingly advanced AI systems. We are excited about the potential for these systems to drive improvements in education, science, provisioning of public services, and more. At the same time, large AI developers themselves warn that their AI systems could pose serious risks, which they have voluntarily committed to addressing. The Report stated that "some risks have unclear but growing evidence...AI-enabled hacking or biological attacks, and loss of control" – the risks that SB 53 aims to address and gather more evidence about. Advanced AI is currently mostly unregulated, and these risks are currently being managed by companies themselves without any requirement that they inform the public about their risk management practices or report serious incidents. SB 53 addresses this much needed gap by implementing four key recommendations from the report.
>
> First, the Report argued that "transparency into the risks associated with foundation models, what mitigations are implemented to address risks, and how the two interrelate is the foundation for understanding how model developers manage risk." SB 53 implements this recommendation as a requirement for large AI developers to write, publish, and follow safety and security protocols to manage the most severe risks. This is in line with voluntary commitments that companies have already made. Rather than prescribe specific technical standards that companies must take, the bill simply requires companies to be transparent about the approaches they are using. Some of the specific required elements of safety protocols, such as a requirement to manage risks related to internal use of AI models and cybersecurity policies, directly mirror recommendations in the Report. Others mirror components of the Stanford Foundation Model Transparency Index, which is cited prominently in the Report.

Second, the Report stated that "transparency into pre-deployment assessments of capabilities and risks, spanning both developer-conducted and externally conducted evaluations, is vital given that these evaluations are early indicators of how models may affect society and may be interpreted (potentially undesirably) as safety assurances." SB 53 accomplishes this with a requirement that large developers publish transparency reports that include the results of their pre-deployment assessments of catastrophic risk. The Report also argues that "transparency into the safety cases used to assess risk provides clarity into how developers justify decisions around model safety," which forms the basis for 22757.12(c)(3).

Third, the Report concluded that "an adverse event reporting system that combines mandatory developer reporting with voluntary user reporting maximally grows the evidence base." SB 53 takes exactly this approach by establishing a tightly defined set of critical safety incidents that AI developers are required to report to the Attorney General. It would also allow members of the public to optionally submit reports.

Finally, the Report recommends strengthening whistleblower protections, pointing out that "actions that may clearly pose a risk and violate company policies...may not violate any existing laws. Therefore, policymakers may consider protections that cover a broader range of activities, which may draw upon notions of 'good faith' reporting on risks found in other domains such as cybersecurity." This recommendation is mirrored in SB 53, which allows employees to report evidence of catastrophic risks as well as violations of SB 53 itself to government authorities with legal protections against retaliation.

SB 53 only applies to the largest AI developers – those training models with more than 10^26 floating point operations (FLOPs). These are companies spending hundreds of millions or billions of dollars to train the most advanced AI models. It would impose no burden on smaller companies and the requirements it imposes on large companies are minimal compared to what companies are already voluntarily doing. The Report argues that "policymakers should ensure that mechanisms are in place to adapt thresholds over time—not only by updating specific threshold values but also by revising or replacing metrics if needed." It also suggests specific criteria that thresholds should be evaluated for. Following this recommendation, SB 53 allows the Attorney General to update the definition of "large developer" through regulation while considering the same factors described in the report. Regardless of any update, the Attorney General must only include "well-resourced large developers at the frontier of artificial intelligence development" in the scoping of the bill. If legislation is needed to cover other developers, the Attorney General is instructed to write a report to the Legislature requesting it.

Finally, SB 53 would also set in motion CalCompute, a public cloud computing cluster for use by academics and startups in California. Computational resources are essential for AI research and CalCompute would make those resources more accessible to California's top universities and startups, helping to catalyze additional research into beneficial applications of AI and supporting, in particular, smaller startups for a healthier innovation ecosystem. This mirrors a similar computing cluster that is already being established in New York state. We support this groundbreaking effort, which would advance and democratize AI research in California.

SB 53 thoughtfully implements the recommendations of the Report by combining a low-burden transparency and reporting regime with a public compute cluster that will broaden

access for AI researchers and startups in California. This is a commonsense approach that will strengthen the AI ecosystem, benefiting both companies and the public interest.

For all these reasons, we respectfully urge your support of this important measure.

Arguments in Opposition
In opposition to the bill, Chamber of Progress argues:

On behalf of the Chamber of Progress, a tech industry association supporting public policies to build a more inclusive society in which all people benefit from technological advances, we respectfully urge you to oppose SB 53, based on its recent amendments.

The definition of "catastrophic risk" remains vague and overreaching

While the amended bill replaces the term "critical risk" with "catastrophic risk," the underlying problem persists. The definition remains overly expansive and ambiguous, capturing a wide array of hypothetical scenarios that may not reflect real-world AI capabilities or threats.

Under Section 22757.11(b), the definition of "catastrophic risk" includes scenarios where a foundation model is "materially likely" to cause harm, potentially due to misuse or malicious inputs. However, this standard is vague, lacks clear and objective thresholds, and leaves room for subjective interpretation by whistleblowers or regulators. In a rapidly evolving field like AI, such ambiguity could unfairly penalize developers who are acting responsibly.

In addition, the inclusion of highly abstract risks, such as the evasion of human control under Section 22757.12(a)(2), creates significant uncertainty. Without clear technical criteria, companies may face liability or investigation based on assumptions about what a model might enable rather than what it has demonstrably done. This uncertainty undermines research and commercial deployment in California and could push critical AI development efforts out of state or abroad.

The $100,000,000 compute cost threshold risks misidentifying frontier AI models

SB 53's use of an arbitrary $100,000,000 compute cost threshold to determine eligibility for protections is an inherently flawed method for identifying frontier AI models. This threshold may result in the overinclusion of developers working on benign systems while potentially excluding smaller models that pose significant real-world risks. It also ignores the constantly changing cost of compute.

A more effective approach would involve a threshold based on model capabilities, deployment context, and specific use cases rather than relying solely on computational costs.

SB 53's extensive safety and security protocols create impractical burdens for AI developers

SB 53 imposes comprehensive safety and security requirements on AI developers, as outlined in Section 22757.12(a), including risk testing, deployment practices, and escalation procedures. While these objectives are important, the bill demands an impractical level of detailed planning and documentation for every conceivable misuse scenario, many of which are speculative or unrealistic.

This exhaustive approach compels developers to allocate significant time and resources toward preparing for hypothetical risks rather than addressing actual, demonstrable harms. For startups and smaller companies, these extensive protocols create a heavy administrative burden that diverts critical resources away from innovation and the timely deployment of beneficial AI technologies.

Additionally, Section 22757.12(c)'s requirement that developers publish detailed transparency reports, before or at the time of deploying a new or substantially modified foundation model, creates significant risks to both competitiveness and operational security.

Although redactions are permitted under subsection (f), the requirement to publish the "character and justification" of redacted material could still inadvertently expose business-sensitive strategies or vulnerabilities. This level of forced transparency goes beyond reasonable accountability and may discourage responsible companies from operating in California. It also creates opportunities for misuse by malicious actors who could exploit disclosed model weaknesses or mitigation gaps.

In fast-moving AI markets, publication of this level of detail erodes a developer's ability to maintain a competitive edge and deters innovation by raising legal and reputational risks associated with even speculative harms.

FISCAL COMMENTS

According to the Assembly Appropriations Committee:

1) Costs (General Fund) to the Department of Justice (DOJ), likely in the low millions of dollars annually, to establish reporting mechanisms, review critical incident reports, conduct investigations, publish reports, and enforce violations. DOJ anticipates costs of approximately $1.1 million in fiscal year 2025-26 and $2 million annually ongoing thereafter for eight staff positions (attorneys, analysts, IT specialists, and legal secretaries) in its Consumer Protection Section and external consultant costs. DOJ reports it is unable to absorb these costs and can implement this bill only with an appropriation of additional funding. DOJ may also incur enforcement costs for violations of the bill's whistleblower protections; the bill does not clearly specify the entity responsible for this enforcement but permits an employee making a whistleblower report to use an existing DOJ whistleblower hotline.

2) Costs (General Fund) to GovOps to establish and operate the CalCompute consortium until January 1, 2027, possibly in the high hundreds of thousands of dollars to low millions of dollars. GovOps estimates total costs of $2.5 million for expert contractors, infrastructure planning, and staffing to manage the project, conduct research, and develop the required report. GovOps was not able to provide a breakdown of these costs but anticipates the workload would be handled by contract workers due to the short timeframe in the bill. Members of the consortium are not entitled to compensation but are entitled to reimbursement for necessary expenses incurred in performing their duties; these costs were not included in GovOps' fiscal estimate but may be in the thousands to low tens of thousands of dollars depending on the activities of the consortium.

3) Possible cost pressures (General Fund) of an unknown but potentially significant amount to the UC to operate CalCompute, should it be established within the UC. State costs may be offset to some extent by private donations, which the bill authorizes the UC to receive to implement CalCompute.

4) Costs (General Fund, Labor and Enforcement Compliance Fund) of an unknown but potentially significant amount to the Labor Commissioner to enforce violations of the bill's whistleblower provisions. Actual costs will depend on the number of violations, the number of actions pursued, and the amount of workload associated with each action.

5) Cost pressures (Trial Court Trust Fund, General Fund) to the courts to adjudicate enforcement actions and whistleblower cases. Actual costs will depend on the number of violations, the number of actions filed, and the amount of court time needed to resolve each case. It generally costs approximately $1,000 to operate a courtroom for one hour. Although courts are not funded on the basis of workload, increased pressure on the Trial Court Trust Fund may create a demand for increased funding for courts from the General Fund. The fiscal year 2025-26 state budget provides $82 million ongoing General Fund to the Trial Court Trust Fund for court operations.

VOTES

SENATE FLOOR: 37-0-3
YES: Allen, Alvarado-Gil, Archuleta, Arreguín, Ashby, Becker, Blakespear, Cabaldon, Caballero, Choi, Cortese, Dahle, Durazo, Gonzalez, Grayson, Grove, Hurtado, Jones, Laird, McGuire, McNerney, Menjivar, Niello, Ochoa Bogh, Padilla, Pérez, Richardson, Rubio, Seyarto, Smallwood-Cuevas, Stern, Strickland, Umberg, Valladares, Wahab, Weber Pierson, Wiener
ABS, ABST OR NV: Cervantes, Limón, Reyes

ASM JUDICIARY: 12-0-0
YES: Kalra, Dixon, Bauer-Kahan, Bryan, Connolly, Harabedian, Macedo, Pacheco, Papan, Sanchez, Stefani, Zbur

ASM PRIVACY AND CONSUMER PROTECTION: 10-0-5
YES: Bauer-Kahan, Dixon, Irwin, Lowenthal, McKinnor, Ortega, Pellerin, Petrie-Norris, Ward, Wilson
ABS, ABST OR NV: Bryan, DeMaio, Macedo, Patterson, Wicks

ASM APPROPRIATIONS: 11-1-3
YES: Wicks, Arambula, Calderon, Caloza, Elhawary, Fong, Mark González, Ahrens, Pacheco, Pellerin, Solache
NO: Tangipa
ABS, ABST OR NV: Sanchez, Dixon, Ta

UPDATED

VERSION: September 02, 2025

CONSULTANT: John Bennett / P. & C.P. / (916) 319-2200 FN: 0001505