SENATE THIRD READING
SB 53 (Wiener)
As Amended September 5, 2025
Majority vote

SUMMARY

Requires 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 the Office of Emergency Management to establish a mechanism by which developers and the public can report critical safety incidents. Provides for whistleblower protections and enforcement by the Attorney General (AG). Requires the Department of Technology to offer guidance to the Legislature on refining the scope over time to reflect technological advances. Provides, upon appropriation, for the creation of a framework to create a public cloud computing cluster.

Major Provisions

1) Requires large frontier developer 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 frontier AI framework 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 frontier 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. *Developers who only reach the compute threshold must publish a higher level summary of the transparency report.*

3) Requires the Office of Emergency Services to establish a mechanism for frontier model developers and the public to report critical safety incidents that have materialized as a result of the use of their models.

4) Requires the Department of Technology to offer guidance to the Legislature on redefining the scope of this bill 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.

Recent amendments do the following:

1) *Omitting the requirement for independent audits starting in 2030.*

2) *Increasing the revenue for a large frontier developer threshold from $100 million to $500 million.*

3) *Excluding foundation models that do not meet the compute threshold.*

4) *Striking the Attorney General's power to issues regulations adjusting the definition of a developer subject to the bill, and replacing it with CDT's annual report making scoping recommendations to the Legislature.*

5) *Narrowing and refining various definitions, including the collapsing of the definition of "dangerous capabilities" into the definition of "catastrophic risk."*

6) *Adding various exemptions, including risks arising from information outputted by the model where the information is in substantially the same form as a publicly available source, risks that would result in loss of the value of equity, and lawful activity of the federal government.*

7) *Recasting safety and security protocols as frontier AI frameworks which only applies to large frontier developers; simplifying disclosure requirements; subjecting frontier developers that make less than $500 million to a less stringent transparency report.*

8) *Reducing the scope of certain categories of critical safety incidents to those that actually result in harm.*

9) *Limiting the prohibition on false or misleading statements by exempting those that were made in good faith and reasonable under the circumstances.*

10) *Reducing the maximum civil penalty from $10 million to $1 million.*

11) *Removal of contractors from whistleblower protections.*

12) *Preemption of local regulation of frontier models.*

13) *Removing risk assessments for models that developers use for internal purposes from public disclosure requirements; summaries of such assessments must be provided to OES and are confidential.*

COMMENTS

In the 2024 legislative session, SB 1047 (Wiener) sought to address concerns surrounding frontier models – the largest and most powerful artificial intelligence (AI) systems – by establishing a regulatory framework intended to prevent the potential catastrophic harms that many experts have warned of. After vetoing the bill, Governor Gavin Newsom convened the Joint California Policy Working Group on AI Frontier Models 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 report's recommendations. Much narrower than its predecessor, SB 53 takes a very light-touch approach that focuses on transparency as the means of ensuring safety and accountability for developers of the most powerful and expensive models – those who harness an extraordinarily high amount of compute power and have over $500 million in annual revenues. Under the bill, such developers must create, implement, and publish a Frontier AI framework – documented technical and organizational protocols to manage, assess, and mitigate catastrophic risks – and a transparency report for each released model. Additionally, developers who only reach the compute threshold must publish a high-level transparency report. The bill does not prescribe any particular standards for these disclosures: 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 Department of Technology (CDT) may offer guidance to the Legislature to redefine the scope of entities subject to the bill to ensure that the bill remains responsive to technological advancements.

The bill also establishes a critical incident reporting mechanism, administered by the Office of Emergency Management, to ensure that severe or high-risk events are tracked and addressed in a timely manner. Incident reports must be made by any frontier model developer within 15 days of the incident, unless the incident presents an imminent threat, in which case the developer must report the incident to law enforcement within 24 hours. The bill also provides whistleblower protections for employees of frontier model developers who report certain risks or noncompliance. Finally, 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.

For a full analysis please see the revised analysis from the Assembly Privacy Consumer Protection Committee hearing on Sept. 11, 2025.

According to the Author
Senate Bill 53 ensures California continues to lead not only on AI innovation, but on responsible practices to help ensure that innovation is safe and secure. It does so by:

1) Requiring covered developers to write, implement, and publish their Frontier AI Framework in redacted form to protect intellectual property;

2) Requiring covered developers to report carefully defined critical safety incidents to the Office of Emergency Services and allowing members of the public to report incidents

3) Prohibiting covered developers from preventing a covered employee from disclosing, or retaliating against covered employee that discloses, that a developer's activities pose a catastrophic risk;

4) Requiring that large frontier developers provide an internal process through which an employee may anonymously disclose information to the developer if the employee believes in good faith that the developer's activities pose a catastrophic risk; and

5) Establishing a process to create a public cloud-computing cluster that will conduct research into the safe and secure deployment of large-scale artificial intelligence (AI) models.

   In doing this, SB 53 allows California to continue to lead in this space and to demonstrate that safety does not stifle success.

Arguments in Support

Anthropic, writes in support:

> As you know, SB 53 would, for the first time, govern powerful AI systems built by frontier AI developers like Anthropic. We've long advocated for thoughtful AI regulation and our support for this bill comes after careful consideration of the lessons learned from California's previous attempt at AI regulation (SB 1047). While we believe that frontier AI safety is ideally addressed at the federal level instead of a patchwork of state regulations, powerful AI advancements won't wait for consensus in Washington.
>
> The measure is also in keeping with direction from Governor Newsom and his Joint California Policy Working Group. The working group endorsed an approach of 'trust but verify', and SB 53 implements this principle through disclosure requirements rather than the prescriptive technical mandates that plagued last year's efforts.
>
> SB 53 would require large companies developing the most powerful AI systems to:
>
> 1) Develop and publish safety frameworks, which describe how they manage, assess, and mitigate catastrophic risks—risks that could foreseeably and materially contribute to a mass casualty incident or substantial monetary damages.
>
> 2) Release public transparency reports summarizing their catastrophic risk assessments and the steps taken to fulfill their respective frameworks before deploying powerful new models.
>
> 3) Report critical safety incidents to the state within 15 days, and even confidentially disclose summaries of any assessments of the potential for catastrophic risk from the use of internally-deployed models.
>
> 4) Provide clear whistleblower protections that cover violations of these requirements as well as substantial dangers to public health/safety from catastrophic risk.
>
> 5) Be publicly accountable for the commitments made in their frameworks or face monetary penalties.
>
> These requirements would formalize practices that Anthropic and many other frontier AI companies already follow. At Anthropic, we publish our Responsible Scaling Policy, detailing how we evaluate and mitigate risks as our models become more capable. We release comprehensive system cards that document model capabilities and limitations. Other frontier labs (Google DeepMind, OpenAI, Microsoft) have adopted similar approaches while vigorously competing at the frontier. Now all covered models will be legally held to this standard. The bill also appropriately focuses on large companies developing the most powerful AI systems, while providing exemptions for smaller companies that are less likely to develop powerful models and should not bear unnecessary regulatory burdens. Of course, no major piece of legislation like SB 53 is perfect, nor do we expect it to be. But what is clear is that SB 53's transparency requirements will have an important impact on frontier AI safety. Without it, labs with increasingly powerful models could face growing incentives to dial back their own safety and disclosure programs in order to compete. But with SB 53, developers can compete while ensuring they remain transparent about AI capabilities that pose risks to public safety, creating a level playing field.

> The question before us all isn't whether we need AI governance—it's whether we'll develop it thoughtfully today or reactively tomorrow. SB 53 offers a solid path toward the former. We commend Senator Wiener and Governor Newsom for their leadership on responsible frontier AI governance, and we encourage the California Legislature to pass SB 53.

Arguments in Opposition
In an oppose-unless-amended position, CalChamber, Computer & Communications Industry Association, and TechNet jointly write:

> [. . .]
>
> We share your goal of ensuring the safe and responsible development of AI and appreciate efforts made in recent amendments to find common ground on how California should approach artificial intelligence models and we appreciate improvements made to the bill over the last several weeks. That being said, there are some issues of concern that remain and wish to flag certain other areas where the bill could be better aligned with the final findings of Governor Newsom's Joint California Policy Working Group on AI Frontier Models, which arose out of his veto of SB 1047 (2024).
>
> *SB 53 should focus on model risk, not developer size—to fully address concerns about powerful models capable of catastrophic risk*
>
> We are concerned about the bill's focus on "large developers" to the exclusion of other developers of models with advanced capabilities that pose risks of catastrophic harm. As amended September 5th, *SB 53* now focuses on models that have a computational threshold of 10^26 floating point operations (or "FLOPs") but only if those models are developed by entities with at least $500m in annual revenues.
>
> Consistent with our position in SB 1047, we maintain that small entities can develop hugely influential and potentially risky models with similar capabilities to the models developed by "large developers", as demonstrated by the Chinese company DeepSeek. As noted above, upon vetoing SB 1047, the Governor commissioned experts in the field to form the Joint California Working Group on AI Frontier Models, which has validated such concerns in their Final Reports, finding that small companies may create powerful models that pose safety risks. By excluding such models here, the bill fails to adequately address the very real risks posed by small but malicious models and imposes significant costs on innovating performant but responsible ones. The Governor's Joint California Policy Working Group on AI Frontier Models cautions against developer-level thresholds stating:
>
>> Generic developer-level thresholds seem to be generally undesirable given the current AI landscape. Since many small entities can develop hugely influential and potentially risky foundation models, as demonstrated by the Chinese company DeepSeek, the use of thresholds based on developer-level properties may inadvertently ignore key players. […] At the same time, these approaches may bring into scope massive, established companies in other industries that are simply exploring the use of AI since thresholds based on properties of companies may not distinguish between the entire business and the AI-specific subset. Therefore, we

> caution against the use of customary developer-level metrics that do not consider the specifics of the AI industry and its associated technology.[1]

SB 53 should make clear that the AI ecosystem includes multiple actors including downstream developers

SB 53 does not account for the complexity of the AI value chain. Models are routinely adapted and fine-tuned by downstream developers in ways that could potentially increase risk. The bill should make clear that a frontier developer's obligations do not extend to models that have been substantially modified by unaffiliated parties, otherwise accountability will be muddled and innovation chilled. We note that whereas the Governor's Work Group report recognized the full AI ecosystem value chain, *SB 53* still needs to fully recognize the roles of not just the original developer of a foundational model but also of those unaffiliated third parties who may modify and/or build on top of a foundation model. The bill should clarify these provisions to reflect the realities of the ecosystem, including downstream developers and open-source models.

SB 53 still raises concerns about protecting trade secrets and sensitive information, including matters of cybersecurity and national security.

We appreciate that amendments were made to change the level of detail required of the AI Safety Framework and changing summaries for transparency reports. However, *SB 53* now requires a large developer only to transmit to the California Office of Emergency Services (CalOES) a summary of any assessment of catastrophic risk resulting from internal use of its frontier models every three months. Not only is this cadence of reporting unnecessary, CalOES will need to take serious steps to protect this information from being accessed by cybercriminals, foreign adversaries, or bad actors. Without ironclad safeguards, these transparency requirements could unintentionally make us less safe. The Joint California Policy Working Group on AI Frontier Models warns against this level of disclosure.

> General details about risks of foundation models can be made public without undermining security, especially if these risks have been demonstrated in other foundation models or AI technologies. Specific details about concrete vulnerabilities should be disclosed carefully, with advanced notice to actors in the supply chain who are able to remediate them prior to broader disclosure.[2]

Requiring developers to justify redactions is less effective than not requiring developers to disclose any information that would include trade secrets, cybersecurity information, or other confidential or proprietary information.

SB 53 unnecessarily re-writes California Whistleblower law for just one industry

As amended, SB 53 rewrites California's already robust whistleblower protections for just one industry. Creating a special, one-off standard for a single sector not only sets a poor precedent but also risks confusion and inconsistency across industries. Current law covers whistleblowing activities associated with AI safety because there is a robust body of existing

[1] Final Report at p.
[2] *Id.* at 30.

law that governs whistleblower protection covering employees who report violations of state/federal laws, rules, or regulations. These laws are intentionally tied to actions that are illegal so there are clear lines of what is considered applicable and understood who gets protection when reporting. These protections cover activities associated with AI without creating unnecessary and confusing new processes in state law.

For example, Labor Code Section 6310 already protects whistleblowers who report unsafe working conditions or work practices. Similarly, federal laws such as the Sarbanes-Oxley Act protect employees who report safety violations or substantial and specific dangers to public health or safety. A brightline threshold is needed for what activity is covered so it is clear when a developer's activities should be reported. For instance, in the field of research and development, innovations are being experimented with in novel contexts where there may be significant disagreement on what actions constitute risk. Thus, the bill mandates that there be an allegation of "*specific and substantial danger to public health or safety resulting from a catastrophic risk*," the inherently subjective nature of these terms leaves room for differing interpretations as to what does or does not meet the threshold.

SB 53 requires steep penalties that are disproportionate for technical errors, inflexible incident reporting requirements, and no right to cure

As amended, *SB 53* imposes a $1 million fine for a possible paperwork error which is excessive and risks punishing good-faith developers for technical mistakes rather than deterring real harm. Penalties should be fair, targeted, and proportionate. As we pointed out in our July 12th letter, *SB 53* requires incident reporting within 15 days but does not provide flexibility for an investigation timeline. Even if 15 days is a reasonable reporting period, requirements should be flexible because all facts may not be known within 15 days of discovery. With respect to enforcement, we again state our view that the bill should grant businesses at least a 60 day right to cure, to ensure that law focuses on compliance and not punishment. In addition, given the highly detailed requirements of the bill as drafted, we think enforcement efforts should be focused on material failures to comply rather than also covering technical paperwork errors.

While we understand your focus on this issue and appreciate the recent amendments have made meaningful improvements to the prior version of the bill, given the immense promise of this technology, we believe that the bill would benefit from a focus on a risk-based framework for all frontier models, additional clarity in responsibilities among actors in the AI value chain, additional safeguards for trade secrets and security, and reasonable timelines, penalties, and enforcement provisions. [. . .]

FISCAL COMMENTS

According to the Assembly Appropriations Committee:

1) Ongoing costs (General Fund) to the Office of Emergency Services (CalOES) to receive summaries of internal risk assessments, review critical safety incident reports, produce annual reports aggregating information about critical safety incident reports, and, if warranted, adopt regulations. Ongoing costs will depend largely on the level of staffing CalOES needs to fulfill these responsibilities and may be in the low millions of dollars annually. CalOES may also incur significant one-time costs for additional workload and IT infrastructure to establish the critical safety incident reporting mechanism.

2) Costs (General Fund) to the Department of Technology (CDT) to conduct assessments, make recommendations regarding specified definitions included in the bill, and submit a report to the Legislature. Actual costs will largely depend on the level of staffing CDT needs to complete these responsibilities, possibly in the hundreds of thousands of dollars annually.

3) Costs (General Fund) to the Department of Justice (DOJ) of an unknown but potentially significant amount to bring civil enforcement actions and produce an annual report aggregating information from whistleblower reports. Actual costs will depend largely on whether the Attorney General pursues enforcement actions, and, if so, the level of additional staffing DOJ needs to handle the related workload. If DOJ hires staff to handle enforcement actions authorized by this bill, costs may be in the low hundreds of thousands of dollars at a minimum.

4) 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 did not 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.

5) 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.

6) Costs (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.

7) 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

ASM PRIVACY AND CONSUMER PROTECTION: 12-1-2
YES: Bauer-Kahan, Dixon, Bryan, Irwin, Lowenthal, McKinnor, Ortega, Pellerin, Petrie-Norris, Ward, Wicks, Wilson
NO: Macedo
ABS, ABST OR NV: DeMaio, Patterson

UPDATED

VERSION: September 5, 2025

CONSULTANT: John Bennett / P. & C.P. / (916) 319-2200 FN: 0002076