Office of Senate Floor Analyses
(916) 651-1520 Fax: (916) 327-4478

UNFINISHED BUSINESS

Bill No: SB 53
Author: Wiener (D), et al.
Amended: 9/5/25 in Assembly
Vote: 21

SENATE GOVERNMENTAL ORG. COMMITTEE: 13-0, 3/25/25
AYES: Padilla, Valladares, Archuleta, Ashby, Blakespear, Cervantes, Hurtado, Jones, Ochoa Bogh, Richardson, Rubio, Wahab, Weber Pierson
NO VOTE RECORDED: Dahle, Smallwood-Cuevas

SENATE JUDICIARY COMMITTEE: 13-0, 4/8/25
AYES: Umberg, Niello, Allen, Arreguín, Ashby, Caballero, Durazo, Laird, Stern, Valladares, Wahab, Weber Pierson, Wiener

SENATE APPROPRIATIONS COMMITTEE: 6-0, 5/23/25
AYES: Caballero, Seyarto, Cabaldon, Grayson, Richardson, Wahab
NO VOTE RECORDED: Dahle

SENATE FLOOR: 37-0, 5/28/25
AYES: 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
NO VOTE RECORDED: Cervantes, Limón, Reyes

ASSEMBLY FLOOR: 45-7, 9/12/25 – Roll call vote not available.

SUBJECT: Artificial intelligence models: large developers

SOURCE: Economic Security California Action
Encode
Secure AI Project

DIGEST: This bill requires large artificial intelligence (AI) developers, as defined, to publish safety frameworks, disclose specified transparency reports, and report critical safety incidents to the Office of Emergency Services (OES), as specified. Additionally, this bill creates enhanced whistleblower protections for employees reporting AI safety violations and establishes a consortium to design a framework for "CalCompute," a public cloud platform to expand safe and equitable AI research, as specified.

Assembly Amendments of 9/5/25 enact the Transparency in Frontier AI Act (TFAIA) and require large AI frontier model developers, as defined, to publish safety frameworks, disclose specified transparency reports, and report critical safety incidents to OES, as specified.

ANALYSIS:

Existing law:

1) Generally regulates AI, including by requiring that a generative AI system or service made publicly available to Californians to use, the developer of the system or service to post on their internet website documentation regarding the data used by the developer to train the AI system or service, as specified.

2) Establishes the California Department of Technology (CDT), within the Government Operations Agency (GovOps), as specified.

3) Establishes the California Cybersecurity Integration Center, within OES, to serve as the central organizing hub of state government's cybersecurity activities and to coordinate information sharing with various entities.

This bill:

1) Requires a large AI frontier developer, as defined, to write, implement, and comply with, and clearly and conspicuously publish on its internet website a frontier AI framework that applies to the large frontier developer's frontier models and describes how the large frontier developer approaches specified related situations.

2) Requires a frontier developer, before deploying a new frontier model or a substantially modified version of an existing frontier model, to clearly and conspicuously publish on its internet website a transparency report, as specified.

3) Requires a large frontier developer to transmit to OES a summary of any assessment of catastrophic risk resulting from internal use of its frontier models every three months or pursuant to another reasonable schedule, as appropriate.

4) Requires OES to establish a mechanism to be used by a frontier developer or a member of the public to report a critical safety incident, as specified.

5) Requires OES, beginning January 1, 2027, and annually thereafter, to produce a report with anonymized and aggregated information about critical safety incidents that have been reviewed by the office since the preceding report, as specified.

6) Requires CDT, on or before January 1, 2027, and annually thereafter, to assess recent evidence and developments relevant to the purposes of this bill and to make recommendations about whether and how to update specified and related definitions in statute.

7) Provides that if a large frontier developer fails to publish or transmit a complaint document as required by this bill, makes a statement in violation of this bill, or fails to comply with its own AI framework shall be subject to a civil penalty in an amount dependent upon the severity of the violation that does not exceed one million dollars ($1,000,000) per violation.

8) Establishes a consortium, within GovOps to develop a framework for the creation of a public cloud computing cluster to be known as "CalCompute," as specified.

9) Prohibits a frontier developer from making, adopting, enforcing, or entering into a rule, regulation, policy, or contract that prevents a covered employee – as defined – from disclosing, or retaliating against a covered employee for disclosing, information to the Attorney General (AG), a federal authority, a person with authority over the covered employee, or another covered employee who has authority to investigate, discover, or correct the reported issue, as specified.

Background

Author Statement. According to the author's office, "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."

Frontier AI Models. "Foundation" or "frontier" models are the largest, most powerful AI systems being built today. Because of their broad capabilities, they have the potential to unlock major breakthroughs in science and medicine, streamline complex processes, and grow the economy. However, they also carry the potential for catastrophic risks. Critics point out that a frontier model might be used to cure disease or, conversely, to engineer a new pandemic. It could automate government functions or, if left unchecked, disrupt critical infrastructure.

Debates around AI safety often center on timing: some argue that regulating too early risks stifling innovation, while others warn that waiting until evidence of harm is undeniable may leave society unable to respond. In 2024, SB 1047 (Wiener) attempted to regulate frontier models directly. It would have required developers to adopt safety and security protocols before training, install shutdown mechanisms, conduct risk assessments, and submit to third-party audits. It also would have barred release of models posing unreasonable catastrophic risks and created a new Board of Frontier Models to oversee compliance. Governor

Newsom vetoed the bill, acknowledging the urgency but stressing that any regulatory framework must be grounded in empirical evidence and able to evolve alongside rapidly advancing technology.

Following the veto, the Governor convened the Joint California Policy Working Group on AI Frontier Models, which issued a final policy report in June 2025. SB 53 attempts to implement some of those recommendations with a narrower, transparency-focused approach. Rather than dictating safety standards, the bill requires the largest frontier developers (those training models with extraordinary compute power and earning over $500 million in annual revenue) to publish both safety protocols and detailed transparency reports for each model. Developers that meet only the compute threshold must release a more limited transparency report. These documents must explain whether and how catastrophic risks (defined as events causing more than 50 deaths or $1 billion in damage) are assessed and mitigated.

SB 53 also sets up a critical incident reporting system through OES, requiring developers to report incidents within 15 days, or within 24 hours if there is an imminent threat. CDT is empowered to recommend updates to the law as AI capabilities evolve.

CalCompute. This bill also creates a consortium to establish a framework for creating "CalCompute," a public cloud cluster within GovOps to expand access to compute for AI research and safety testing, and extends whistleblower protections to employees who report risks or violations. Cluster computing is a type of computing where multiple computers are connected to work together as a single system. Computing clusters typically consist of servers, workstations, and personal computers that communicate over a local area network or a wide area network.

Whistleblower Protections. This bill adds safeguards from the Working Group Report: capability thresholds must be disclosed in protocols and reports, mitigation steps must be documented when thresholds are exceeded, and OES and the Attorney General must issue annual, anonymized summaries of critical incidents and whistleblower disclosures.

Whistleblower protections within companies developing AI foundation models can be crucial given the extraordinary scale and impact of the technology. These protections serve as an essential safety valve in an industry where the potential consequences of unethical or dangerous development practices can affect millions, if not billions, of people.

These models increasingly power critical infrastructure across healthcare, finance, employment, education, and government services. When employees with direct knowledge of risks, harms, or unethical practices cannot safely speak up, dangerous systems may be deployed without proper safeguards or public awareness. Whistleblowers often represent the last line of defense when corporate incentives prioritize growth, profit, or competitive advantage over public welfare.

Related/Prior Legislation

SB 1047 (Wiener, 2024) would have required developers of powerful artificial intelligence models and those providing the computing power to train such models to put appropriate safeguards and policies into place to prevent critical harms. This bill would have also established a state entity to oversee the development of these models and called for the creation of a consortium to develop a framework for a public cloud computing cluster. (Vetoed by Governor Newsom)

FISCAL EFFECT: Appropriation: No Fiscal Com.: Yes Local: No

According to the Assembly Appropriations Committee, 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 cots 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.

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.

Possible cost pressures (General Fund) of an unknown but potentially significant amount to the University of California (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.

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.

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.

SUPPORT: (Verified 9/9/2025)

Economic Security California Action (co-source)
Encode (co-source)
Secure AI Project (co-source)
AI for Animals
AI Futures Project
AI Lab Watch
AI Policy Tracker
All Girls Allowed
Apart Research
Association for Long Term Existence and Resilience
Berkeley Existential Risk Initiative
California Democratic Party
California Federation of Labor Unions, AFL-CIO
California Initiative for Technology & Democracy
Center for AI and Digital Policy
Center for AI Policy

Center for Human-Compatible AI
Center for Youth and AI
Children's Advocacy Institute, University of San Diego School of Law
Common Sense Media
Depict.ai
District Council of Iron Workers of the State of California and Vicinity
EarningsStream LLC
Elicit
Eon Systems
Existential Risk Observatory
Frontlines Foundation
Future of Life Institute
Indivisible California Statestrong
Little Hoover Commission
Momentum
Nonlinear
Oakland Privacy
Omidyar Network
Redwood Research
Safe AI Future
Scorecard
Secure Ai Future
SEIU Califonia
Tech Oversight California
TechEquity Action
The Brandes Lab At NYU
The Midas Project
The Signals Network
Transparency Coalition.ai
Trevi Digital Assets Fund
University of California
Youth Leadership Institute

OPPOSITION: (Verified 9/9/25)

Business Software Alliance
California Chamber of Commerce
Computer & Communications Industry Association
Consumer Technology Association
Insights Association
Los Angeles County Business Federation

Silicon Valley Leadership Group
TechNet

ARGUMENTS IN SUPPORT: In support of the bill, Secure AI Project, co-sponsors of the bill, alongside a coalition of technology equity advocacy groups write that:

> 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."

ARGUMENTS IN OPPOSITION: In opposition to this bill, the 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.

Prepared by: Brian Duke / G.O. / (916) 651-1530
9/12/25 22:23:23

****** END ******