AMENDED IN ASSEMBLY SEPTEMBER 5, 2025

AMENDED IN ASSEMBLY SEPTEMBER 2, 2025

AMENDED IN ASSEMBLY JULY 17, 2025

AMENDED IN ASSEMBLY JULY 8, 2025

AMENDED IN SENATE MAY 23, 2025

AMENDED IN SENATE MARCH 27, 2025

AMENDED IN SENATE FEBRUARY 27, 2025

SENATE BILL No. 53

Introduced by Senator Wiener
(Coauthor: Senator Rubio)

January 7, 2025

An act to add Chapter 25.1 (commencing with Section 22757.10) to Division 8 of the Business and Professions Code, to add Section 11546.8 to the Government Code, and to add Chapter 5.1 (commencing with Section 1107) to Part 3 of Division 2 of the Labor Code, relating to artificial intelligence.

LEGISLATIVE COUNSEL'S DIGEST

SB 53, as amended, Wiener. Artificial intelligence models: large developers.

(1) Existing law generally regulates artificial intelligence, including by requiring, on or before January 1, 2026, and before each time thereafter, that a generative artificial intelligence system or service, or a substantial modification to a generative artificial intelligence system or service, released on or after January 1, 2022, is made publicly available to Californians for use, the developer of the system or service

to post on the developer's internet website documentation regarding the data used by the developer to train the generative artificial intelligence system or service, as prescribed.

This bill would enact the Transparency in Frontier Artificial Intelligence Act (TFAIA) that would, among other things related to ensuring the safety of a foundation ~~model~~ *model, as defined,* developed by a ~~large~~ *frontier* developer, require a large *frontier* developer to write, implement, and clearly and conspicuously publish on its internet website a ~~safety and security protocol that describes in specific detail,~~ *frontier AI framework that applies to the large frontier developer's frontier models and describes how the large frontier developer approaches,* among other things, ~~the testing procedures that the large developer uses to assess catastrophic risks from its foundation models, as specified.~~ *incorporating national standards, international standards, and industry-consensus best practices into its frontier AI framework.* The TFAIA would ~~define "foundation model" to mean an artificial intelligence model that is trained on a broad data set, designed for generality of output, and adaptable to a wide range of distinctive tasks.~~ *also require a large frontier developer to transmit to the Office of Emergency Services a summary of any assessment of catastrophic risk, as defined, resulting from internal use of its frontier models, as specified.* The TFAIA would require the ~~Attorney General~~ *Office of Emergency Services* to establish a mechanism to be used by a ~~large~~ *frontier* developer or a member of the public to report, as prescribed, a critical safety incident, as ~~defined, and would authorize the Attorney General to periodically update the definition of "foundation model," as specified.~~ *defined, and would also require the Office of Emergency Services to establish a mechanism to be used by a large frontier developer to confidentially submit summaries of any assessments of the potential for catastrophic risk resulting from internal use of its frontier models, as prescribed.*

The TFAIA would exempt from the California Public Records Act a report of a critical safety incident submitted to the ~~Attorney General~~ *Office of Emergency Services, a report of assessments of catastrophic risk from internet use,* and ~~an~~ *a covered* employee report made pursuant to the whistleblower protections described below.

The ~~bill~~ *TFAIA* would impose a civil penalty for noncompliance with the TFAIA to be enforced by the Attorney General, as prescribed.

(2) Existing law establishes the Department of Technology within the Government Operations Agency. Existing law requires the

department to conduct, in coordination with other interagency bodies as it deems appropriate, a comprehensive inventory of all high-risk automated decision systems that have been proposed for use, development, or procurement by, or are being used, developed, or procured by, any state agency.

This bill would establish within 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 artificial intelligence that is safe, ethical, equitable, and sustainable by, among other things, fostering research and innovation that benefits the public, as prescribed. The bill would require the Government Operations Agency to, on or before January 1, 2027, submit a report from the consortium to the Legislature with that framework and would dissolve the consortium upon submission of that report. The bill would make those provisions operative only upon an appropriation in a budget act, or other measure, for its purposes.

(3) Existing law prohibits employers and their agents from making, adopting, or enforcing a rule, regulation, or policy preventing an employee from disclosing information to certain entities or from providing information to, or testifying before, any public body conducting an investigation, hearing, or inquiry if the employee has reasonable cause to believe that the information discloses a violation of a law, as specified, and prohibits retaliation against an employee for, among other things, exercising these rights.

This bill would, among other things related to protecting whistleblowers working with foundation models, prohibit a ~~large~~ *frontier* developer from making, adopting, enforcing, or entering into a rule, regulation, policy, or contract that prevents ~~an employee~~ *a covered employee, as defined,* from disclosing, or retaliates against ~~an~~ *a covered* employee for disclosing, information to the Attorney General, 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, if the *covered* employee has reasonable cause to believe that the information discloses that the ~~large~~ *frontier* developer's activities pose a *specific and substantial danger to the public health or safety resulting from a* catastrophic risk or that the ~~large~~ *frontier* developer has violated the TFAIA.

This bill would require a large *frontier* developer to provide a certain internal process through which ~~an~~ *a covered* employee may

anonymously disclose information to the large *frontier* developer if the *covered* employee believes in good faith that the information indicates that the large *frontier* developer's activities present a *specific and substantial danger to the public health or safety resulting from a* catastrophic ~~risk,~~ *risk* or that the *large frontier* developer ~~has made a false or misleading statement about its management of catastrophic risk, as prescribed.~~ *violated the TFAIA.* The bill would specify provisions particular to the enforcement of those whistleblower protections and would authorize attorney's fees to a plaintiff who brings a successful action for a violation.

This bill would preempt any rule, regulation, code, ordinance, or other law adopted by a city, county, city and county, municipality, or local agency on or after January 1, 2025, specifically related to the regulation of frontier developers with respect to their management of catastrophic risk.

Existing constitutional provisions require that a statute that limits the right of access to the meetings of public bodies or the writings of public officials and agencies be adopted with findings demonstrating the interest protected by the limitation and the need for protecting that interest.

This bill would make legislative findings to that effect.

Vote: majority. Appropriation: no. Fiscal committee: yes. State-mandated local program: no.

The people of the State of California do enact as follows:

SECTION 1. The Legislature finds and declares all of the following:

(a) California is leading the world in artificial intelligence innovation and research through companies large and small and through the state's remarkable public and private universities.

(b) Artificial intelligence, including new advances in foundation models, has the potential to catalyze innovation and the rapid development of a wide range of benefits for Californians and the California economy, including advances in medicine, wildfire forecasting and prevention, and climate science, and to push the bounds of human creativity and capacity.

(c) The Joint California Policy Working Group on AI Frontier Models has recommended sound principles for policy in artificial intelligence.

(d) Targeted interventions to support effective artificial intelligence governance should balance the technology's benefits and *the potential for* material risks.

~~(e) Artificial intelligence developers have already voluntarily committed to creating safety and security protocols and releasing the results of risk assessments.~~

~~(f)~~

(e) In building a robust and transparent evidence environment, policymakers can align incentives to simultaneously protect consumers, leverage industry expertise, and recognize leading safety practices.

~~(g) When~~

(f) As industry actors conduct internal research on their technologies' impacts, ~~a significant information asymmetry can develop between those with privileged access to data and the broader public.~~ *public trust in these technologies would significantly benefit from access to information regarding, and increased awareness of, frontier AI capabilities.*

~~(h)~~

(g) Greater ~~transparency, given current information deficits,~~ *transparency* can *also* advance accountability, competition, and public trust.

~~(i)~~

(h) Whistleblower protections and public-facing information sharing are key instruments to increase transparency.

~~(j) Adverse event~~

(i) Incident reporting systems enable monitoring of the post-deployment impacts of artificial intelligence.

~~(k) There is growing evidence that, unless~~

(j) Unless they are developed with careful diligence and reasonable precaution, *there is concern that* advanced artificial intelligence systems could *have capabilities that* pose catastrophic risks from both malicious uses and malfunctions, including artificial intelligence-enabled hacking, biological attacks, and loss of control.

~~(l)~~

(k) With the frontier of artificial intelligence rapidly evolving, there is a need for legislation to track the frontier of artificial intelligence research and alert policymakers and the public to ~~the~~ *serious* risks and harms from the very most advanced artificial

intelligence systems, while avoiding burdening smaller companies behind the frontier.

(l) While the major artificial intelligence developers have already voluntarily established the creation, use, and publication of frontier AI frameworks as an industry best practice, not all developers are providing reporting that is consistent and sufficient to ensure necessary transparency and protection of the public. Mandatory, standardized, and objective reporting by frontier developers is required to provide the government and the public with timely and accurate information.

(m) Timely reporting of critical safety incidents to the government is essential to ensure that public authorities are promptly informed of ongoing and emerging risks to public safety. This reporting enables the government to monitor, assess, and respond effectively in the event that advanced capabilities emerge in frontier artificial intelligence models that may pose a threat to the public.

~~(m)~~

(n) In the future, foundation models developed by smaller companies or that are behind the frontier may pose significant catastrophic risk, and additional legislation may be needed at that time.

(o) The recent release of the Governor's California Report on Frontier AI Policy and testimony from legislative hearings on artificial intelligence before the Legislature reflect the advances in AI model capabilities that could pose potential catastrophic risk in frontier artificial intelligence, which this act aims to address.

~~(n)~~

(p) It is the intent of the Legislature to create more transparency, but collective safety will depend in part on ~~large~~ *frontier* developers taking due care in their development and deployment of ~~foundation~~ *frontier* models proportional to the scale of the foreseeable risks.

SEC. 2. Chapter 25.1 (commencing with Section 22757.10) is added to Division 8 of the Business and Professions Code, to read:

CHAPTER 25.1. TRANSPARENCY IN FRONTIER ARTIFICIAL INTELLIGENCE ACT

22757.10. This chapter shall be known as the Transparency in Frontier Artificial Intelligence Act.

22757.11. For purposes of this chapter:

(a) "Affiliate" means a person controlling, controlled by, or under common control with a specified person, directly or indirectly, through one or more intermediaries.

~~(a)~~

(b) "Artificial intelligence model" means an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.

~~(b)~~

(c) (1) "Catastrophic risk" means a foreseeable and material risk that a ~~large~~ *frontier* developer's development, storage, use, or deployment of a ~~foundation~~ *frontier* model will materially contribute to the death of, or serious injury to, more than 50 people or more than one billion dollars ($1,000,000,000) in damage to, or loss of, property arising from a single ~~incident, scheme, or course of conduct~~ *incident* involving a ~~dangerous capability.~~ *frontier model doing any of the following:*

(A) Providing expert-level assistance in the creation or release of a chemical, biological, radiological, or nuclear weapon.

(B) Engaging in conduct with no meaningful human oversight, intervention, or supervision that is either a cyberattack or, if the conduct had been committed by a human, would constitute the crime of murder, assault, extortion, or theft, including theft by false pretense.

(C) Evading the control of its frontier developer or user.

(2) "Catastrophic risk" does not include a foreseeable and material risk from any of the following:

(A) Information that a frontier model outputs if the information is otherwise publicly accessible in a substantially similar form from a source other than a foundation model.

(B) Lawful activity of the federal government.

(C) Harm caused by a frontier model in combination with other software if the frontier model did not materially contribute to the harm.

~~(c)~~

(d) "Critical safety incident" means any of the following:

(1) Unauthorized access to, modification of, or exfiltration of, the model weights of a ~~foundation model.~~ *frontier model that results in death or bodily injury.*

(2) Harm resulting from the materialization of a catastrophic risk.

(3) Loss of control of a ~~foundation~~ *frontier* model causing ~~death, bodily injury, or damage to, or loss of, property.~~ *death or bodily injury.*

(4) A ~~foundation~~ *frontier* model that uses deceptive techniques against the ~~large~~ *frontier* developer to subvert the controls or monitoring of its ~~large~~ *frontier* developer outside of the context of an evaluation designed to elicit this ~~behavior.~~ *behavior and in a manner that demonstrates materially increased catastrophic risk.*

~~(5) Attaining a dangerous capability or catastrophic risk threshold, as defined in the large developer's safety and security protocol pursuant to paragraph (3) of subdivision (a) of Section 22757.12, for the first time.~~

~~(d) "Dangerous capability" means the capacity of a foundation model to do any of the following:~~

~~(1) Provide expert-level assistance in the creation or release of a chemical, biological, radiological, or nuclear weapon.~~

~~(2) Conduct or assist in a cyberattack.~~

~~(3) Engage in conduct, with limited human intervention, that would, if committed by a human, constitute the crime of murder, assault, extortion, or theft, including theft by false pretense.~~

~~(4) Evade the control of its large developer or user.~~

(e) (1) "Deploy" means to make a ~~foundation~~ *frontier* model available to a third party for use, modification, copying, or combination with other software.

(2) "Deploy" does not include making a ~~foundation~~ *frontier* model available to a third party for the primary purpose of developing or evaluating the ~~foundation~~ *frontier* model.

(f) "Foundation model" means an artificial intelligence model that is all of the following:

(1) Trained on a broad data set.
(2) Designed for generality of output.
(3) Adaptable to a wide range of distinctive tasks.

(g) "Frontier AI framework" means documented technical and organizational protocols to manage, assess, and mitigate catastrophic risks.

(h) "Frontier developer" means a person who has trained, or initiated the training of, a frontier model, with respect to which the person has used, or intends to use, at least as much computing power to train the frontier model as would meet the technical specifications found in subdivision (i).

(i) (1) "Frontier model" means a foundation model that was trained using a quantity of computing power greater than 10^26 integer or floating-point operations.

(2) The quantity of computing power described in paragraph (1) shall include computing for the original training run and for any subsequent fine-tuning, reinforcement learning, or other material modifications the developer applies to a preceding foundation model.

~~(g)~~

(j) "Large *frontier* developer" means ~~either of the following:~~ *a frontier developer that together with its affiliates collectively had annual gross revenues in excess of five hundred million dollars ($500,000,000) in the preceding calendar year.*

~~(1) (A) Before January 1, 2027, "large developer" means a person who meets both of the following criteria:~~

~~(i) (I) The person has trained, or initiated the training of, at least one foundation model using a quantity of computing power greater than 10^26 integer or floating-point operations.~~

~~(II) The quantity of computing power described in subclause (I) shall include computing for the original training run and any subsequent fine-tuning, reinforcement learning, or other material modifications to a preceding foundation model.~~

~~(ii) The person had annual gross revenues in excess of five hundred million dollars ($500,000,000) in the preceding calendar year.~~

~~(2) (A) Except as provided in subparagraph (B), on and after January 1, 2027, "large developer" has the meaning defined by a regulation adopted by the Attorney General pursuant to Section 22757.15.~~

~~(B) If the Attorney General does not adopt a regulation described in subparagraph (A) by January 1, 2027, the definition in paragraph (1) shall be operative until the regulation is adopted.~~

~~(h)~~

(k) "Model weight" means a numerical parameter in a ~~foundation~~ *frontier* model that is adjusted through training and that helps determine how inputs are transformed into outputs.

~~(i)~~

(l) "Property" means tangible or intangible property.

~~(j) "Safety and security protocol" means documented technical and organizational protocols to manage, assess, and mitigate catastrophic risks.~~

22757.12. (a) A large *frontier* developer shall write, implement, comply with, and clearly and conspicuously publish on its internet website a ~~safety and security protocol~~ *frontier AI framework* that ~~describes in specific detail~~ *applies to the large frontier developer's frontier models and describes how the large frontier developer approaches* all of the following:

~~(1) How, if at all, the large developer excludes certain foundation models from being covered by its safety and security protocol because those foundation models are incapable of posing material catastrophic risks and which foundation models, if any, have been included.~~

~~(2) The testing procedures that the large developer uses to assess catastrophic risks from its foundation models, including risk resulting from malfunctions, misuse, and foundation models evading the control of the large developer or user.~~

~~(3) (A) Thresholds used by the large developer to identify and assess whether a foundation model has attained a dangerous capability or poses a catastrophic risk.~~

~~(B) How the large developer will assess whether a threshold described in subparagraph (A) has been attained, which may include multiple tiered thresholds for different dangerous capabilities or for catastrophic risk.~~

~~(C) The actions the large developer will take if each threshold is attained.~~

~~(4) The mitigations that a large developer takes to reduce a catastrophic risk and how the large developer assesses the effectiveness of those mitigations.~~

~~(5) The degree to which the large developer's assessments of catastrophic risk and dangerous capabilities and the effectiveness of catastrophic risk mitigations are reproducible by external entities.~~

~~(6) The extent to which, and how, a large developer will use third parties to assess catastrophic risks and dangerous capabilities and the effectiveness of mitigations of catastrophic risk.~~

~~(7) The large developer's cybersecurity practices and how the large developer secures unreleased model weights from unauthorized modification or transfer by internal or external parties.~~

~~(8) To the extent that the foundation model is controlled by the large developer, the procedures the large developer will use to monitor critical safety incidents and the steps that a large developer would take to respond to a critical safety incident, including, but not limited to, whether the large developer has the ability to promptly shut down copies of foundation models owned and controlled by the large developer, who the large developer will notify, and the timeline on which the large developer would take these steps.~~

~~(9) The testing procedures that the large developer will use to assess and manage a catastrophic risk or dangerous capability resulting from the internal use of its foundation models, including risks resulting from a foundation model circumventing oversight mechanisms, and the schedule, specified in days, by which the large developer will report these assessments pursuant to subdivision (d).~~

~~(10) How the developer determines when its foundation models are substantially modified enough to conduct additional assessments and publish a transparency report pursuant to subdivision (c).~~

(1) Incorporating national standards, international standards, and industry-consensus best practices into its frontier AI framework.

(2) Defining and assessing thresholds used by the large frontier developer to identify and assess whether a frontier model has capabilities that could pose a catastrophic risk, which may include multiple-tiered thresholds.

(3) Applying mitigations to address the potential for catastrophic risks based on the results of assessments undertaken pursuant to paragraph (2).

(4) Reviewing assessments and adequacy of mitigations as part of the decision to deploy a frontier model or use it extensively internally.

(5) Using third parties to assess the potential for catastrophic risks and the effectiveness of mitigations of catastrophic risks.

(6) Revisiting and updating the frontier AI framework, including any criteria that trigger updates and how the large frontier developer determines when its frontier models are substantially modified enough to require disclosures pursuant to subdivision (c).

(7) Cybersecurity practices to secure unreleased model weights from unauthorized modification or transfer by internal or external parties.

(8) Identifying and responding to critical safety incidents.

(9) Instituting internal governance practices to ensure implementation of these processes.

(10) Assessing and managing catastrophic risk resulting from the internal use of its frontier models, including risks resulting from a frontier model circumventing oversight mechanisms.

(b) (1) A large frontier developer shall review and, as appropriate, update its frontier AI framework at least once per year.

~~(b)~~

(2) If a large *frontier* developer makes a material modification to its ~~safety and security protocol,~~ *frontier AI framework,* the large *frontier* developer shall clearly and conspicuously publish the modified ~~protocol~~ *frontier AI framework* and a justification for that modification within 30 days.

(c) *(1)* Before, or concurrently with, deploying a new ~~foundation~~ *frontier* model or a substantially modified version of an existing ~~foundation model that is covered by the large developer's safety and security protocol,~~ *frontier model,* a ~~large~~ *frontier* developer shall clearly and conspicuously publish on its internet website a transparency report containing ~~both~~ *all* of the following:

~~(1) (A) The results of any risk assessment or risk mitigation assessment conducted by the large developer or a third party that contracts with a large developer pursuant to its safety and security protocol, why the information gathered by the large developer or third party leads to the stated results, and the steps taken to address any identified risks.~~

~~(B) A large developer shall disclose the time and extent of predeployment access provided to any third party described in subparagraph (A), whether or not the third party was independent, and the nature of any constraints the large developer placed on the assessment or on the third party's ability to disclose information about its assessment to the public or to government officials.~~

~~(C) Whether a catastrophic risk threshold or dangerous capability threshold has been attained for the foundation model and any actions taken as a result.~~

~~(2) (A) The reasoning behind the large developer's decision to deploy the foundation model, the process by which the large developer arrived at that decision, and any limitations in the assessments that the large developer used to make that decision.~~

~~(B) A large developer may reuse an answer previously provided under subparagraph (A) if the rationale in question has not materially changed for the new deployment.~~

(A) The internet website of the frontier developer.

(B) A mechanism that enables a natural person to communicate with the frontier developer.

(C) The release date of the frontier model.

(D) The languages supported by the frontier model.

(E) The modalities of output supported by the frontier model.

(F) The intended uses of the frontier model.

(G) Any generally applicable restrictions or conditions on uses of the frontier model.

(2) Before, or concurrently with, deploying a new frontier model or a substantially modified version of an existing frontier model, a large frontier developer shall include in the transparency report required by paragraph (1) summaries of all of the following:

(A) Assessments of catastrophic risks from the frontier model conducted pursuant to the large frontier developer's frontier AI framework.

(B) The results of those assessments.

(C) The extent to which third-party evaluators were involved.

(D) Other steps taken to fulfill the requirements of the frontier AI framework with respect to the frontier model.

(3) A frontier developer that publishes the information described in paragraph (1) or (2) as part of a larger document, including a system card or model card, shall be deemed in compliance with the applicable paragraph.

(4) A frontier developer is encouraged, but not required, to make disclosures described in this subdivision that are consistent with, or superior to, industry best practices.

(d) A large *frontier* developer shall ~~clearly and conspicuously publish on its internet website~~ *transmit to the Office of Emergency Services a summary of* any assessment of catastrophic risk ~~or dangerous capabilities~~ resulting from internal use of its ~~foundation~~ *frontier* models ~~pursuant to the schedule the developer specifies in its safety and security protocol.~~ *every three months or pursuant to another reasonable schedule specified by the large frontier developer and communicated in writing to the Office of Emergency Services with written updates, as appropriate.*

(e) *(1) (A)* A ~~large~~ *frontier* developer shall not make a materially false or misleading statement about catastrophic risk from its ~~foundation models,~~ *frontier models or* its management of catastrophic ~~risk, or its implementation of or compliance with its safety and security protocol.~~ *risk.*

(B) A large frontier developer shall not make a materially false or misleading statement about its implementation of, or compliance with, its frontier AI framework.

(2) This subdivision does not apply to a statement that was made in good faith and was reasonable under the circumstances.

(f) (1) When a ~~large~~ *frontier* developer publishes documents to comply with this section, the ~~large~~ *frontier* developer may make redactions to those documents that are necessary to protect the ~~large~~ *frontier* developer's trade secrets, the ~~large~~ *frontier* developer's cybersecurity, public safety, or the national security of the United States or to comply with any federal or state law.

(2) If a ~~large~~ *frontier* developer redacts information in a document pursuant to this subdivision, the ~~large~~ *frontier* developer shall describe the character and justification of the redaction in any published version of the document to the extent permitted by the concerns that justify redaction and shall retain the unredacted information for five years.

22757.13. (a) The ~~Attorney General~~ *Office of Emergency Services* shall establish a mechanism to be used by a ~~large~~ *frontier* developer or a member of the public to report a critical safety incident that includes all of the following:

(1) The date of the critical safety incident.

(2) The reasons the incident qualifies as a critical safety incident.

(3) A short and plain statement describing the critical safety incident.

(4) Whether the incident was associated with internal use of a frontier model.

(b) (1) The Office of Emergency Services shall establish a mechanism to be used by a large frontier developer to confidentially submit summaries of any assessments of the potential for catastrophic risk resulting from internal use of its frontier models.

(2) The Office of Emergency Services shall take all necessary precautions to limit access to any reports related to internal use of frontier models to only personnel with a specific need to know the information and to protect the reports from unauthorized access.

~~(b)~~

(c) (1) Subject to paragraph (2), a ~~large~~ *frontier* developer shall report any critical safety incident pertaining to one or more of its ~~foundation~~ *frontier* models to the ~~Attorney General~~ *Office of Emergency Services* within 15 days of discovering the critical safety incident.

(2) If a ~~large~~ *frontier* developer discovers that a critical safety incident poses an imminent risk of death or serious physical injury, the ~~large~~ *frontier* developer shall disclose that incident within 24 hours to an authority, including any law enforcement agency or public safety agency with jurisdiction, that is appropriate based on the nature of that incident and as required by law.

(3) A frontier developer that discovers information about a critical safety incident after filing the initial report required by this subdivision may file an amended report.

(4) A frontier developer is encouraged, but not required, to report critical safety incidents pertaining to foundation models that are not frontier models.

~~(c)~~

(d) The ~~Attorney General~~ *Office of Emergency Services* shall review critical safety incident reports submitted by ~~large~~ *frontier* developers and may review reports submitted by members of the public.

~~(d)~~

(e) (1) The Attorney General *or the Office of Emergency Services* may transmit reports of critical safety incidents and reports

from *covered* employees made pursuant to Chapter 5.1 (commencing with Section 1107) of Part 3 of Division 2 of the Labor Code to the Legislature, the Governor, the federal government, or appropriate state agencies.

(2) The Attorney General or the Office of Emergency Services shall strongly consider any risks related to trade secrets, public safety, cybersecurity of a frontier developer, or national security when transmitting reports.

~~(e)~~

(f) A report of a critical safety incident submitted to the ~~Attorney General~~ *Office of Emergency Services* pursuant to this ~~section~~ *section, a report of assessments of catastrophic risk from internal use pursuant to Section 22757.12,* and ~~an~~ *a covered* employee report made pursuant to Chapter 5.1 (commencing with Section 1107) of Part 3 of Division 2 of the Labor Code are exempt from the California Public Records Act (Division 10 (commencing with Section 7920.000) of Title 1 of the Government Code).

~~(f)~~

(g) (1) Beginning January 1, 2027, and annually thereafter, the ~~Attorney General~~ *Office of Emergency Services* shall produce a report with anonymized and aggregated information about critical safety incidents ~~and reports from employees made pursuant to Chapter 5.1 (commencing with Section 1107) of Part 3 of Division 2 of the Labor Code~~ that have been reviewed by the ~~Attorney General~~ *Office of Emergency Services* since the preceding report.

(2) The ~~Attorney General~~ *Office of Emergency Services* shall not include information in a report pursuant to this subdivision that would compromise the trade secrets or cybersecurity of a ~~large~~ *frontier* developer, public safety, or the national security of the United States or that would be prohibited by any federal or state law.

(3) The ~~Attorney General~~ *Office of Emergency Services* shall transmit a report pursuant to this subdivision to the Legislature, pursuant to Section 9795, and to the Governor.

(h) The Office of Emergency Services may adopt regulations designating one or more federal laws, regulations, or guidance documents that meet all of the following conditions for the purposes of subdivision (i):

(1) (A) The law, regulation, or guidance document imposes or states standards or requirements for critical safety incident

reporting that are substantially equivalent to, or stricter than, those required by this section.

(B) The law, regulation, or guidance document described in subparagraph (A) does not need to require critical safety incident reporting to the State of California.

(2) The law, regulation, or guidance document is intended to assess, detect, or mitigate the catastrophic risk.

(i) (1) A frontier developer that intends to comply with this section by complying with the requirements of, or meeting the standards stated by, a federal law, regulation, or guidance document designated pursuant to subdivision (h) shall declare its intent to do so to the Office of Emergency Services.

(2) After a frontier developer has declared its intent pursuant to paragraph (1), both of the following apply:

(A) The frontier developer shall be deemed in compliance with this section to the extent that the frontier developer meets the standards of, or complies with the requirements imposed or stated by, the designated federal law, regulation, or guidance document until the frontier developer declares the revocation of that intent to the Office of Emergency Services or the Office of Emergency Services revokes a relevant regulation pursuant to subdivision (j).

(B) The failure by a frontier developer to meet the standards of, or comply with the requirements stated by, the federal law, regulation, or guidance document designated pursuant to subdivision (h) shall constitute a violation of this chapter.

(j) The Office of Emergency Services shall revoke a regulation adopted under subdivision (h) if the requirements of subdivision (h) are no longer met.

22757.14. (a) On or before January 1, 2027, and annually thereafter, the ~~Attorney General may adopt regulations~~ *Department of Technology shall assess recent evidence and developments relevant to the purposes of this chapter and shall make recommendations about whether and how* to update ~~the definition of a "large developer"~~ *any of the following definitions* for the purposes of this chapter to ensure that ~~it~~ *they* accurately ~~reflects~~ *reflect* technological developments, scientific literature, and widely accepted national and international ~~standards and applies to well-resourced large developers at the frontier of artificial intelligence development.~~ *standards:*

(1) "Frontier model" so that it applies to foundation models at the frontier of artificial intelligence development.

(2) "Frontier developer" so that it applies to developers of frontier models who are themselves at the frontier of artificial intelligence development.

(3) "Large frontier developer" so that it applies to well-resourced frontier developers.

(b) In ~~developing regulations~~ *making recommendations* pursuant to this section, the ~~Attorney General~~ *Department of Technology* shall take into account all of the following:

(1) Similar thresholds used in international standards or federal law, guidance, or regulations for the management of catastrophic ~~risk.~~ *risk and shall align with a definition adopted in a federal law or regulation to the extent that it is consistent with the purposes of this chapter.*

(2) Input from stakeholders, including academics, industry, the open-source community, and governmental entities.

(3) The extent to which a person will be able to determine, before beginning to train or deploy a foundation model, whether that person will be subject to the ~~regulations~~ *definition* as a *frontier developer or as a* large *frontier* developer with an aim toward allowing earlier determinations if possible.

(4) The complexity of determining whether a person *or foundation model* is covered, with an aim toward allowing simpler determinations if possible.

(5) The external verifiability of determining whether a person *or foundation model* is covered, with an aim toward definitions that are verifiable by parties other than the ~~large~~ *frontier* developer.

~~(c) If the Attorney General determines that less well-resourced developers, or developers significantly behind the frontier of artificial intelligence, may create substantial catastrophic risk, the Attorney General shall promptly submit a report to the Legislature, pursuant to Section 9795, with a proposal for managing this source of catastrophic risk but shall not include those developers within the definition of "large developer" without authorization in subsequently enacted legislation.~~

(c) Upon developing recommendations pursuant to this section, the Department of Technology shall submit a report to the Legislature, pursuant to Section 9795 of the Government Code, with those recommendations.

(d) (1) Beginning January 1, 2027, and annually thereafter, the Attorney General shall produce a report with anonymized and aggregated information about reports from covered employees made pursuant to Chapter 5.1 (commencing with Section 1107) of Part 3 of Division 2 of the Labor Code that have been reviewed by the Attorney General since the preceding report.

(2) The Attorney General shall not include information in a report pursuant to this subdivision that would compromise the trade secrets or cybersecurity of a frontier developer, confidentiality of a covered employee, public safety, or the national security of the United States or that would be prohibited by any federal or state law.

(3) The Attorney General shall transmit a report pursuant to this subdivision to the Legislature, pursuant to Section 9795 of the Government Code, and to the Governor.

22757.15. (a) A large *frontier* developer ~~who violates this chapter~~ *that fails to publish or transmit a compliant document required to be published or transmitted under this chapter, makes a statement in violation of subdivision (e) of Section 22757.12, fails to report an incident as required by Section 22757.13, or fails to comply with its own frontier AI framework* shall be subject to a civil penalty ~~subject to all of the following:~~ *in an amount dependent upon the severity of the violation that does not exceed one million dollars ($1,000,000) per violation.*

~~(1) (A) For an unknowing violation that does not create a material risk of death, serious physical injury, or a catastrophic risk, a large developer shall be subject to a civil penalty in an amount not to exceed ten thousand dollars ($10,000).~~

~~(B) (i) If the violation is a first violation subject to a civil penalty pursuant to subparagraph (A), a large developer shall be provided with a 30-day period to cure the violation after notification by the Attorney General. If the violation is cured within that period, a civil penalty shall not be imposed for that violation.~~

~~(ii) For the purposes of this subparagraph, a violation involving the missing or late publication or submission of a document is cured when the developer publishes or submits that document, even if the publication or submission occurs after the statutory deadline.~~

~~(2) For a knowing violation that does not create a material risk of death, serious physical injury, or a catastrophic risk or an~~

~~unknowing violation that creates a material risk of death, serious physical injury, or a catastrophic risk, a large developer shall be subject to a civil penalty in an amount not to exceed one hundred thousand dollars ($100,000).~~

~~(3) For a knowing violation that creates a material risk of death, serious physical injury, or a catastrophic risk, a large developer shall be subject to a civil penalty in an amount not to exceed one million dollars ($1,000,000) for a violation that is the large developer's first such violation and in an amount not exceeding ten million dollars ($10,000,000) for any subsequent violation.~~

(b) A civil penalty described in this section shall be recovered in a civil action brought only by the Attorney General.

22757.16. The loss of value of equity does not count as damage to or loss of property for the purposes of this chapter.

SEC. 3. Section 11546.8 is added to the Government Code, to read:

11546.8. (a) There is hereby established within the Government Operations Agency a consortium that shall develop, pursuant to this section, a framework for the creation of a public cloud computing cluster to be known as "CalCompute."

(b) The consortium shall develop a framework for the creation of CalCompute that advances the development and deployment of artificial intelligence that is safe, ethical, equitable, and sustainable by doing, at a minimum, both of the following:

(1) Fostering research and innovation that benefits the public.

(2) Enabling equitable innovation by expanding access to computational resources.

(c) The consortium shall make reasonable efforts to ensure that CalCompute is established within the University of California to the extent possible.

(d) CalCompute shall include, but not be limited to, all of the following:

(1) A fully owned and hosted cloud platform.

(2) Necessary human expertise to operate and maintain the platform.

(3) Necessary human expertise to support, train, and facilitate the use of CalCompute.

(e) The consortium shall operate in accordance with all relevant labor and workforce laws and standards.

(f) (1) On or before January 1, 2027, the Government Operations Agency shall submit, pursuant to Section 9795, a report from the consortium to the Legislature with the framework developed pursuant to subdivision (b) for the creation and operation of CalCompute.

(2) The report required by this subdivision shall include all of the following elements:

(A) A landscape analysis of California's current public, private, and nonprofit cloud computing platform infrastructure.

(B) An analysis of the cost to the state to build and maintain CalCompute and recommendations for potential funding sources.

(C) Recommendations for the governance structure and ongoing operation of CalCompute.

(D) Recommendations for the parameters for use of CalCompute, including, but not limited to, a process for determining which users and projects will be supported by CalCompute.

(E) An analysis of the state's technology workforce and recommendations for equitable pathways to strengthen the workforce, including the role of CalCompute.

(F) A detailed description of any proposed partnerships, contracts, or licensing agreements with nongovernmental entities, including, but not limited to, technology-based companies, that demonstrates compliance with the requirements of subdivisions (c) and (d).

(G) Recommendations regarding how the creation and ongoing management of CalCompute can prioritize the use of the current public sector workforce.

(g) The consortium shall, consistent with state constitutional law, consist of 14 members as follows:

(1) Four representatives of the University of California and other public and private academic research institutions and national laboratories appointed by the Secretary of Government Operations.

(2) Three representatives of impacted workforce labor organizations appointed by the Speaker of the Assembly.

(3) Three representatives of stakeholder groups with relevant expertise and experience, including, but not limited to, ethicists, consumer rights advocates, and other public interest advocates appointed by the Senate Rules Committee.

(4) Four experts in technology and artificial intelligence to provide technical assistance appointed by the Secretary of Government Operations.

(h) The members of the consortium shall serve without compensation, but shall be reimbursed for all necessary expenses actually incurred in the performance of their duties.

(i) The consortium shall be dissolved upon submission of the report required by paragraph (1) of subdivision (f) to the Legislature.

(j) If CalCompute is established within the University of California, the University of California may receive private donations for the purposes of implementing CalCompute.

(k) This section shall become operative only upon an appropriation in a budget act, or other measure, for the purposes of this section.

SEC. 4. Chapter 5.1 (commencing with Section 1107) is added to Part 3 of Division 2 of the Labor Code, to read:

CHAPTER 5.1. WHISTLEBLOWER PROTECTIONS: CATASTROPHIC RISKS IN AI FOUNDATION MODELS

1107. For purposes of this chapter:

~~(a) "Artificial intelligence model" means an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.~~

~~(b) "Catastrophic risk" has the meaning defined in Section 22757.11 of the Business and Professions Code.~~

(a) (1) "Catastrophic risk" means a foreseeable and material risk that a frontier developer's development, storage, use, or deployment of a foundation model will materially contribute to the death of, or serious injury to, more than 50 people or more than one billion dollars ($1,000,000,000) in damage to, or loss of, property arising from a single incident involving a foundation model doing any of the following:

(A) Providing expert-level assistance in the creation or release of a chemical, biological, radiological, or nuclear weapon.

(B) Engaging in conduct with no meaningful human oversight, intervention, or supervision that is either a cyberattack or, if

committed by a human, would constitute the crime of murder, assault, extortion, or theft, including theft by false pretense.

(C) Evading the control of its frontier developer or user.

(2) "Catastrophic risk" does not include a foreseeable and material risk from any of the following:

(A) Information that a foundation model outputs if the information is otherwise publicly accessible in a substantially similar form from a source other than a foundation model.

(B) Lawful activity of the federal government.

(C) Harm caused by a foundation model in combination with other software where the foundation model did not materially contribute to the harm.

(b) "Covered employee" means an employee responsible for assessing, managing, or addressing risk of critical safety incidents.

(c) "Critical safety incident" means any of the following:

(1) Unauthorized access to, modification of, or exfiltration of the model weights of a foundation model that results in death, bodily injury, or damage to, or loss of, property.

(2) Harm resulting from the materialization of a catastrophic risk.

(3) Loss of control of a foundation model causing death or bodily injury.

(4) A foundation model that uses deceptive techniques against the frontier developer to subvert the controls or monitoring of its frontier developer outside of the context of an evaluation designed to elicit this behavior and in a manner that demonstrates materially increased catastrophic risk.

(d) "Foundation model" has the meaning defined in Section 22757.11 of the Business and Professions Code.

(e) "Frontier developer" has the meaning defined in Section 22757.11 of the Business and Professions Code.

~~(c)~~

(f) "Large *frontier* developer" has the meaning defined in Section 22757.11 of the Business and Professions Code.

~~(d) "Employee" means a person who performs services for an employer, including both of the following:~~

~~(1) A contractor, subcontractor, or an unpaid advisor involved with assessing, managing, or addressing catastrophic risk, including all of the following:~~

~~(A) An independent contractor.~~

~~(B) A freelance worker.~~
~~(C) A person employed by a labor contractor.~~
~~(D) A board member.~~
~~(2) Corporate officers.~~
~~(e) "Foundation model" has the same meaning as defined in Section 22757.11 of the Business and Professions Code.~~
~~(f) "Labor contractor" means an individual or entity that supplies, either with or without a contract, a client employer with workers to perform labor within the client employer's usual course of business.~~

1107.1. (a) A ~~large~~ *frontier* developer shall not make, adopt, enforce, or enter into a rule, regulation, policy, or contract that prevents ~~an~~ *a covered* employee from disclosing, or retaliates against ~~an~~ *a covered* employee for disclosing, information to the Attorney General, 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, if the *covered* employee has reasonable cause to believe that the information discloses either of the following:

(1) The ~~large~~ *frontier* developer's activities pose a *specific and substantial danger to the public health or safety resulting from a* catastrophic risk.

(2) The ~~large~~ *frontier* developer has violated Chapter 25.1 (commencing with Section 22757.10) of Division 8 of the Business and Professions Code.

(b) A ~~large~~ *frontier* developer shall not enter into a contract that prevents ~~an~~ *a covered* employee from making a disclosure protected under Section 1102.5.

~~(c) A large developer shall not make, adopt, enforce, or enter into a rule, regulation, policy, or contract that would prevent an organization or entity that provides goods or services to the large developer related to the assessment, management, or addressing of catastrophic risk, or an employee of that organization or entity, from disclosing information to the Attorney General, a federal authority, or the developer if the organization, entity, or individual has reasonable cause to believe that the information discloses either of the following:~~
~~(1) The large developer's activities pose a catastrophic risk.~~

~~(2) The large developer has violated Chapter 25.1 (commencing with Section 22757.10) of Division 8 of the Business and Professions Code.~~

~~(d) An~~

(c) A covered employee may use the hotline described in Section 1102.7 to make reports described in subdivision (a).

~~(e)~~

(d) A ~~large~~ *frontier* developer shall provide a clear notice to all *covered* employees of their rights and responsibilities under this section, including by doing either of the following:

(1) At all times posting and displaying within any workplace maintained by the ~~large~~ *frontier* developer a notice to all *covered* employees of their rights under this section, ensuring that any new *covered* employee receives equivalent notice, and ensuring that any *covered* employee who works remotely periodically receives an equivalent notice.

(2) At least once each year, providing written notice to each *covered* employee of the *covered* employee's rights under this section and ensuring that the notice is received and acknowledged by all of those *covered* employees.

~~(f)~~

(e) (1) A large *frontier* developer shall provide a reasonable internal process through which ~~an~~ *a covered* employee may anonymously disclose information to the large *frontier* developer if the *covered* employee believes in good faith that the information indicates that the large *frontier* developer's activities present *a specific and substantial danger to the public health or safety resulting from* a catastrophic risk or that the large *frontier* developer violated Chapter 25.1 (commencing with Section 22757.10) of Division 8 of the Business and Professions Code, including a monthly update to the person who made the disclosure regarding the status of the large *frontier* developer's investigation of the disclosure and the actions taken by the large *frontier* developer in response to the disclosure.

(2) (A) Except as provided in subparagraph (B), the disclosures and responses of the process required by this subdivision shall be shared with officers and directors of the large *frontier* developer at least once each quarter.

(B) If ~~an~~ *a covered* employee has alleged wrongdoing by an officer or director of the large *frontier* developer in a disclosure

or response, subparagraph (A) shall not apply with respect to that officer or director.

~~(g)~~

(f) The court is authorized to award reasonable attorney's fees to a plaintiff who brings a successful action for a violation of this section.

~~(h)~~

(g) In a civil action brought pursuant to this section, once it has been demonstrated by a preponderance of the evidence that an activity proscribed by this section was a contributing factor in the alleged prohibited action against the *covered* employee, the ~~large~~ *frontier* developer shall have the burden of proof to demonstrate by clear and convincing evidence that the alleged action would have occurred for legitimate, independent reasons even if the *covered* employee had not engaged in activities protected by this section.

~~(i)~~

(h) (1) In a civil action or administrative proceeding brought pursuant to this section, ~~an~~ *a covered* employee may petition the superior court in any county wherein the violation in question is alleged to have occurred, or wherein the person resides or transacts business, for appropriate temporary or preliminary injunctive relief.

(2) Upon the filing of the petition for injunctive relief, the petitioner shall cause notice thereof to be served upon the person, and thereupon the court shall have jurisdiction to grant temporary injunctive relief as the court deems just and proper.

(3) In addition to any harm resulting directly from a violation of this section, the court shall consider the chilling effect on other *covered* employees asserting their rights under this section in determining whether temporary injunctive relief is just and proper.

(4) Appropriate injunctive relief shall be issued on a showing that reasonable cause exists to believe a violation has occurred.

(5) An order authorizing temporary injunctive relief shall remain in effect until an administrative or judicial determination or citation has been issued, or until the completion of a review pursuant to subdivision (b) of Section 98.74, whichever is longer, or at a certain time set by the court. Thereafter, a preliminary or permanent injunction may be issued if it is shown to be just and proper. Any temporary injunctive relief shall not prohibit a ~~large~~ *frontier*

developer from disciplining or terminating ~~an~~ *a covered* employee for conduct that is unrelated to the claim of the retaliation.

~~(j)~~

(i) Notwithstanding Section 916 of the Code of Civil Procedure, injunctive relief granted pursuant to this section shall not be stayed pending appeal.

~~(k)~~

(j) (1) This section does not impair or limit the applicability of Section ~~1102.5.~~ *1102.5, including with respect to the rights of employees who are not covered employees to report violations of this chapter or Chapter 25.1 (commencing with Section 22757.10) of Division 8 of the Business and Professions Code.*

(2) The remedies provided by this section are cumulative to each other and the remedies or penalties available under all other laws of this state.

1107.2. The loss of value of equity does not count as damage to or loss of property for the purposes of this chapter.

SEC. 5. (a) The provisions of this act are severable. If any provision of this act or its application is held invalid, that invalidity shall not affect other provisions or applications that can be given effect without the invalid provision or application.

(b) This act shall be liberally construed to effectuate its purposes.

(c) The duties and obligations imposed by this act are cumulative with any other duties or obligations imposed under other law and shall not be construed to relieve any party from any duties or obligations imposed under other law and do not limit any rights or remedies under existing law.

(d) This act shall not apply to the extent that it strictly conflicts with the terms of a contract between a federal government entity and a ~~large~~ *frontier* developer.

(e) This act shall not apply to the extent that it is preempted by federal law.

(f) This act preempts any rule, regulation, code, ordinance, or other law adopted by a city, county, city and county, municipality, or local agency on or after January 1, 2025, specifically related to the regulation of frontier developers with respect to their management of catastrophic risk.

SEC. 6. The Legislature finds and declares that Section 2 of this act, which adds Chapter 25.1 (commencing with Section 22757.10) to Division 8 of the Business and Professions Code,

imposes a limitation on the public's right of access to the meetings of public bodies or the writings of public officials and agencies within the meaning of Section 3 of Article I of the California Constitution. Pursuant to that constitutional provision, the Legislature makes the following findings to demonstrate the interest protected by this limitation and the need for protecting that interest:

Information in critical safety incident ~~reports~~ *reports, assessments of risks from internal use,* and reports from *covered* employees may contain information that could threaten public safety or compromise the response to an incident if disclosed to the public.

O