SB-1047 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act. (2023-2024)

Date Published: 09/03/2024 09:00 PM

ENROLLED SEPTEMBER 03, 2024

PASSED IN SENATE AUGUST 29, 2024

PASSED IN ASSEMBLY AUGUST 28, 2024

AMENDED IN ASSEMBLY AUGUST 22, 2024

AMENDED IN ASSEMBLY AUGUST 19, 2024

AMENDED IN ASSEMBLY JULY 03, 2024

AMENDED IN ASSEMBLY JUNE 20, 2024

AMENDED IN ASSEMBLY JUNE 05, 2024

AMENDED IN SENATE MAY 16, 2024

AMENDED IN SENATE APRIL 30, 2024

AMENDED IN SENATE APRIL 16, 2024

AMENDED IN SENATE APRIL 08, 2024

AMENDED IN SENATE MARCH 20, 2024

CALIFORNIA LEGISLATURE— 2023–2024 REGULAR SESSION

SENATE BILL NO. 1047

Introduced by Senator Wiener
(Coauthors: Senators Roth, Rubio, and Stern)

February 07, 2024

An act to add Chapter 22.6 (commencing with Section 22602) to Division 8 of the Business and Professions Code, and to add Sections 11547.6 and 11547.6.1 to the Government Code, relating to artificial intelligence.

LEGISLATIVE COUNSEL'S DIGEST

SB 1047, Wiener. Safe and Secure Innovation for Frontier Artificial Intelligence Models Act.

Existing law requires the Secretary of Government Operations to develop a coordinated plan to, among other things, investigate the feasibility of, and obstacles to, developing standards and technologies for state departments to determine digital content provenance. For the purpose of informing that coordinated plan, existing law requires the secretary to evaluate, among other things, the impact of the proliferation of deepfakes, defined to mean audio or visual content that has been generated or manipulated by artificial intelligence that would falsely appear to be authentic or truthful and that features depictions of people appearing to say or do things they did not say or do without their consent, on state government, California-based businesses, and residents of the state.

This bill would enact the Safe and Secure Innovation for Frontier Artificial Intelligence Models Act to, among other things, require that a developer, before beginning to initially train a covered model, as defined, comply with various requirements, including implementing the capability to promptly enact a full shutdown, as defined, and implement a written and separate safety and security protocol, as specified. The bill would require a developer to retain an unredacted copy of the safety and security protocol for as long as the covered model is made available for commercial, public, or foreseeably public use plus 5 years, including records and dates of any updates or revisions and would require a developer to grant to the Attorney General access to the unredacted safety and security protocol. The bill would prohibit a developer from using a covered model or covered model derivative for a purpose not exclusively related to the training or reasonable evaluation of the covered model or compliance with state or federal law or making a covered model or a covered model derivative available for commercial or public, or foreseeably public, use, if there is an unreasonable risk that the covered model or covered model derivative will cause or materially enable a critical harm, as defined. The bill would require a developer, beginning January 1, 2026, to annually retain a third-party auditor to perform an independent audit of compliance with those provisions, as prescribed. The bill would require the auditor to produce an audit report, as prescribed, and would require a developer to retain an unredacted copy of the audit report for as long as the covered model is made available for commercial, public, or foreseeably public use plus 5 years. The bill would require a developer to grant to the Attorney General access to the unredacted auditor's report upon request. The bill would exempt from disclosure under the California Public Records Act the safety and security protocol and the auditor's report described above.

This bill would require a developer of a covered model to submit to the Attorney General a statement of compliance with these provisions, as specified. The bill would also require a developer of a covered model to report each artificial intelligence safety incident affecting the covered model or any covered model derivative controlled by the developer to the Attorney General, as prescribed.

This bill would require a person that operates a computing cluster, as defined, to implement written policies and procedures to do certain things when a customer utilizes compute resources that would be sufficient to train a covered model, including assess whether a prospective customer intends to utilize the computing cluster to train a covered model.

This bill would authorize the Attorney General to bring a civil action, as provided. The bill would also provide for whistleblower protections, including by prohibiting a developer of a covered model or a contractor or subcontractor of the developer from preventing an employee from disclosing information, or retaliating against an employee for disclosing information, to the Attorney General or Labor Commissioner if the employee has reasonable cause to believe the information indicates the developer is out of compliance with certain requirements or that the covered model poses an unreasonable risk of critical harm.

This bill would create the Board of Frontier Models within the Government Operations Agency, independent of the Department of Technology, and provide for the board's membership. The bill would require the Government Operations Agency to, on or before January 1, 2027, and annually thereafter, issue regulations to, among other things, update the definition of a "covered model," as provided, and would require the regulations to be approved by the board before taking effect.

This bill would establish 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 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, on or before January 1, 2026, require the Government Operations Agency to submit a report from the consortium to the Legislature with that framework. The bill would make those provisions operative only upon an appropriation in a budget act for its purposes.

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 Local Program: no

THE PEOPLE OF THE STATE OF CALIFORNIA DO ENACT AS FOLLOWS:

SECTION 1. This act shall be known, and may be cited, as the Safe and Secure Innovation for Frontier Artificial Intelligence Models Act.

SEC. 2. 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, as well as through our remarkable public and private universities.

(b) Artificial intelligence, including new advances in generative artificial intelligence, 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) If not properly subject to human controls, future development in artificial intelligence may also have the potential to be used to create novel threats to public safety and security, including by enabling the creation and the proliferation of weapons of mass destruction, such as biological, chemical, and nuclear weapons, as well as weapons with cyber-offensive capabilities.

(d) The state government has an essential role to play in ensuring that California recognizes the benefits of this technology while avoiding the most severe risks, as well as to ensure that artificial intelligence innovation and access to compute is accessible to academic researchers and startups, in addition to large companies.

SEC. 3. Chapter 22.6 (commencing with Section 22602) is added to Division 8 of the Business and Professions Code, to read:

CHAPTER 22.6. Safe and Secure Innovation for Frontier Artificial Intelligence Models

22602. As used in this chapter:

(a) "Advanced persistent threat" means an adversary with sophisticated levels of expertise and significant resources that allow it, through the use of multiple different attack vectors, including, but not limited to, cyber, physical, and deception, to generate opportunities to achieve its objectives that are typically to establish and extend its presence within the information technology infrastructure of organizations for purposes of exfiltrating information or to undermine or impede critical aspects of a mission, program, or organization or place itself in a position to do so in the future.

(b) "Artificial intelligence" 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.

(c) "Artificial intelligence safety incident" means an incident that demonstrably increases the risk of a critical harm occurring by means of any of the following:

(1) A covered model or covered model derivative autonomously engaging in behavior other than at the request of a user.

(2) Theft, misappropriation, malicious use, inadvertent release, unauthorized access, or escape of the model weights of a covered model or covered model derivative.

(3) The critical failure of technical or administrative controls, including controls limiting the ability to modify a covered model or covered model derivative.

(4) Unauthorized use of a covered model or covered model derivative to cause or materially enable critical harm.

(d) "Computing cluster" means a set of machines transitively connected by data center networking of over 100

gigabits per second that has a theoretical maximum computing capacity of at least 10^20 integer or floating-point operations per second and can be used for training artificial intelligence.

(e) (1) "Covered model" means either of the following:

(A) Before January 1, 2027, "covered model" means either of the following:

(i) An artificial intelligence model trained using a quantity of computing power greater than 10^26 integer or floating-point operations, the cost of which exceeds one hundred million dollars ($100,000,000) when calculated using the average market prices of cloud compute at the start of training as reasonably assessed by the developer.

(ii) An artificial intelligence model created by fine-tuning a covered model using a quantity of computing power equal to or greater than three times 10^25 integer or floating-point operations, the cost of which, as reasonably assessed by the developer, exceeds ten million dollars ($10,000,000) if calculated using the average market price of cloud compute at the start of fine-tuning.

(B) (i) Except as provided in clause (ii), on and after January 1, 2027, "covered model" means any of the following:

(I) An artificial intelligence model trained using a quantity of computing power determined by the Government Operations Agency pursuant to Section 11547.6 of the Government Code, the cost of which exceeds one hundred million dollars ($100,000,000) when calculated using the average market price of cloud compute at the start of training as reasonably assessed by the developer.

(II) An artificial intelligence model created by fine-tuning a covered model using a quantity of computing power that exceeds a threshold determined by the Government Operations Agency, the cost of which, as reasonably assessed by the developer, exceeds ten million dollars ($10,000,000) if calculated using the average market price of cloud compute at the start of fine-tuning.

(ii) If the Government Operations Agency does not adopt a regulation governing subclauses (I) and (II) of clause (i) before January 1, 2027, the definition of "covered model" in subparagraph (A) shall be operative until the regulation is adopted.

(2) On and after January 1, 2026, the dollar amount in this subdivision shall be adjusted annually for inflation to the nearest one hundred dollars ($100) based on the change in the annual California Consumer Price Index for All Urban Consumers published by the Department of Industrial Relations for the most recent annual period ending on December 31 preceding the adjustment.

(f) "Covered model derivative" means any of the following:

(1) An unmodified copy of a covered model.

(2) A copy of a covered model that has been subjected to post-training modifications unrelated to fine-tuning.

(3) (A) (i) Before January 1, 2027, a copy of a covered model that has been fine-tuned using a quantity of computing power not exceeding three times 10^25 integer or floating point operations, the cost of which, as reasonably assessed by the developer, exceeds ten million dollars ($10,000,000) if calculated using the average market price of cloud compute at the start of fine-tuning.

(ii) On and after January 1, 2027, a copy of a covered model that has been fine-tuned using a quantity of computing power not exceeding a threshold determined by the Government Operations Agency, the cost of which, as reasonably assessed by the developer, exceeds ten million dollars ($10,000,000) if calculated using the average market price of cloud compute at the start of fine-tuning.

(B) If the Government Operations Agency does not adopt a regulation governing clause (ii) of subparagraph (A) by January 1, 2027, the quantity of computing power specified in clause (i) of subparagraph (A) shall continue to apply until the regulation is adopted.

(4) A copy of a covered model that has been combined with other software.

(g) (1) "Critical harm" means any of the following harms caused or materially enabled by a covered model or covered model derivative:

(A) The creation or use of a chemical, biological, radiological, or nuclear weapon in a manner that results in

mass casualties.

(B) Mass casualties or at least five hundred million dollars ($500,000,000) of damage resulting from cyberattacks on critical infrastructure by a model conducting, or providing precise instructions for conducting, a cyberattack or series of cyberattacks on critical infrastructure.

(C) Mass casualties or at least five hundred million dollars ($500,000,000) of damage resulting from an artificial intelligence model engaging in conduct that does both of the following:

(i) Acts with limited human oversight, intervention, or supervision.

(ii) Results in death, great bodily injury, property damage, or property loss, and would, if committed by a human, constitute a crime specified in the Penal Code that requires intent, recklessness, or gross negligence, or the solicitation or aiding and abetting of such a crime.

(D) Other grave harms to public safety and security that are of comparable severity to the harms described in subparagraphs (A) to (C), inclusive.

(2) "Critical harm" does not include any of the following:

(A) Harms caused or materially enabled by information that a covered model or covered model derivative outputs if the information is otherwise reasonably publicly accessible by an ordinary person from sources other than a covered model or covered model derivative.

(B) Harms caused or materially enabled by a covered model combined with other software, including other models, if the covered model did not materially contribute to the other software's ability to cause or materially enable the harm.

(C) Harms that are not caused or materially enabled by the developer's creation, storage, use, or release of a covered model or covered model derivative.

(3) On and after January 1, 2026, the dollar amounts in this subdivision shall be adjusted annually for inflation to the nearest one hundred dollars ($100) based on the change in the annual California Consumer Price Index for All Urban Consumers published by the Department of Industrial Relations for the most recent annual period ending on December 31 preceding the adjustment.

(h) "Critical infrastructure" means assets, systems, and networks, whether physical or virtual, the incapacitation or destruction of which would have a debilitating effect on physical security, economic security, public health, or safety in the state.

(i) "Developer" means a person that performs the initial training of a covered model either by training a model using a sufficient quantity of computing power and cost, or by fine-tuning an existing covered model or covered model derivative using a quantity of computing power and cost greater than the amount specified in subdivision (e).

(j) "Fine-tuning" means adjusting the model weights of a trained covered model or covered model derivative by exposing it to additional data.

(k) "Full shutdown" means the cessation of operation of all of the following:

(1) The training of a covered model.

(2) A covered model controlled by a developer.

(3) All covered model derivatives controlled by a developer.

(l) "Model weight" means a numerical parameter in an artificial intelligence model that is adjusted through training and that helps determine how inputs are transformed into outputs.

(m) "Person" means an individual, proprietorship, firm, partnership, joint venture, syndicate, business trust, company, corporation, limited liability company, association, committee, or any other nongovernmental organization or group of persons acting in concert.

(n) "Post-training modification" means modifying the capabilities of a covered model or covered model derivative by any means, including, but not limited to, fine-tuning, providing the model with access to tools or data,

removing safeguards against hazardous misuse or misbehavior of the model, or combining the model with, or integrating it into, other software.

(o) "Safety and security protocol" means documented technical and organizational protocols that meet both of the following criteria:

(1) The protocols are used to manage the risks of developing and operating covered models and covered model derivatives across their life cycle, including risks posed by causing or enabling or potentially causing or enabling the creation of covered model derivatives.

(2) The protocols specify that compliance with the protocols is required in order to train, operate, possess, and provide external access to the developer's covered model and covered model derivatives.

22603. (a) Before beginning to initially train a covered model, the developer shall do all of the following:

(1) Implement reasonable administrative, technical, and physical cybersecurity protections to prevent unauthorized access to, misuse of, or unsafe post-training modifications of, the covered model and all covered model derivatives controlled by the developer that are appropriate in light of the risks associated with the covered model, including from advanced persistent threats or other sophisticated actors.

(2) (A) Implement the capability to promptly enact a full shutdown.

(B) When enacting a full shutdown, the developer shall take into account, as appropriate, the risk that a shutdown of the covered model, or particular covered model derivatives, could cause disruptions to critical infrastructure.

(3) Implement a written and separate safety and security protocol that does all of the following:

(A) Specifies protections and procedures that, if successfully implemented, would successfully comply with the developer's duty to take reasonable care to avoid producing a covered model or covered model derivative that poses an unreasonable risk of causing or materially enabling a critical harm.

(B) States compliance requirements in an objective manner and with sufficient detail and specificity to allow the developer or a third party to readily ascertain whether the requirements of the safety and security protocol have been followed.

(C) Identifies a testing procedure, which takes safeguards into account as appropriate, that takes reasonable care to evaluate if both of the following are true:

(i) A covered model poses an unreasonable risk of causing or enabling a critical harm.

(ii) Covered model derivatives pose an unreasonable risk of causing or enabling a critical harm.

(D) Describes in detail how the testing procedure assesses the risks associated with post-training modifications.

(E) Describes in detail how the testing procedure addresses the possibility that a covered model or covered model derivative can be used to make post-training modifications or create another covered model in a manner that may cause or materially enable a critical harm.

(F) Describes in detail how the developer will fulfill their obligations under this chapter.

(G) Describes in detail how the developer intends to implement the safeguards and requirements referenced in this section.

(H) Describes in detail the conditions under which a developer would enact a full shutdown.

(I) Describes in detail the procedure by which the safety and security protocol may be modified.

(4) Ensure that the safety and security protocol is implemented as written, including by designating senior personnel to be responsible for ensuring compliance by employees and contractors working on a covered model, or any covered model derivatives controlled by the developer, monitoring and reporting on implementation.

(5) Retain an unredacted copy of the safety and security protocol for as long as the covered model is made

available for commercial, public, or foreseeably public use plus five years, including records and dates of any updates or revisions.

(6) Conduct an annual review of the safety and security protocol to account for any changes to the capabilities of the covered model and industry best practices and, if necessary, make modifications to the policy.

(7) (A) (i) Conspicuously publish a copy of the redacted safety and security protocol and transmit a copy of the redacted safety and security protocol to the Attorney General.

(ii) A redaction in the safety and security protocol may be made only if the redaction is reasonably necessary to protect any of the following:

(I) Public safety.

(II) Trade secrets, as defined in Section 3426.1 of the Civil Code.

(III) Confidential information pursuant to state and federal law.

(B) The developer shall grant to the Attorney General access to the unredacted safety and security protocol upon request.

(C) A safety and security protocol disclosed to the Attorney General pursuant to this paragraph is exempt from the California Public Records Act (Division 10 (commencing with Section 7920.000) of Title 1 of the Government Code).

(D) If the safety and security protocol is materially modified, conspicuously publish and transmit to the Attorney General an updated redacted copy within 30 days of the modification.

(8) Take reasonable care to implement other appropriate measures to prevent covered models and covered model derivatives from posing unreasonable risks of causing or materially enabling critical harms.

(b) Before using a covered model or covered model derivative for a purpose not exclusively related to the training or reasonable evaluation of the covered model or compliance with state or federal law or before making a covered model or covered model derivative available for commercial or public, or foreseeably public, use, the developer of a covered model shall do all of the following:

(1) Assess whether the covered model is reasonably capable of causing or materially enabling a critical harm.

(2) Record, as and when reasonably possible, and retain for as long as the covered model is made available for commercial, public, or foreseeably public use plus five years information on the specific tests and test results used in the assessment pursuant to paragraph (1) that provides sufficient detail for third parties to replicate the testing procedure.

(3) Take reasonable care to implement appropriate safeguards to prevent the covered model and covered model derivatives from causing or materially enabling a critical harm.

(4) Take reasonable care to ensure, to the extent reasonably possible, that the covered model's actions and the actions of covered model derivatives, as well as critical harms resulting from their actions, can be accurately and reliably attributed to them.

(c) A developer shall not use a covered model or covered model derivative for a purpose not exclusively related to the training or reasonable evaluation of the covered model or compliance with state or federal law or make a covered model or a covered model derivative available for commercial or public, or foreseeably public, use, if there is an unreasonable risk that the covered model or covered model derivative will cause or materially enable a critical harm.

(d) A developer of a covered model shall annually reevaluate the procedures, policies, protections, capabilities, and safeguards implemented pursuant to this section.

(e) (1) Beginning January 1, 2026, a developer of a covered model shall annually retain a third-party auditor that conducts audits consistent with best practices for auditors to perform an independent audit of compliance with the requirements of this section.

(2) An auditor shall conduct audits consistent with regulations issued by the Government Operations Agency pursuant to subdivision (d) of Section 11547.6 of the Government Code.

(3) The auditor shall be granted access to unredacted materials as necessary to comply with the auditor's obligations under this subdivision.

(4) The auditor shall produce an audit report including all of the following:

(A) A detailed assessment of the developer's steps to comply with the requirements of this section.

(B) If applicable, any identified instances of noncompliance with the requirements of this section, and any recommendations for how the developer can improve its policies and processes for ensuring compliance with the requirements of this section.

(C) A detailed assessment of the developer's internal controls, including its designation and empowerment of senior personnel responsible for ensuring compliance by the developer, its employees, and its contractors.

(D) The signature of the lead auditor certifying the results of the auditor.

(5) The developer shall retain an unredacted copy of the audit report for as long as the covered model is made available for commercial, public, or foreseeably public use plus five years.

(6) (A) (i) The developer shall conspicuously publish a redacted copy of the auditor's report and transmit to the Attorney General a copy of the redacted auditor's report.

(ii) A redaction in the auditor's report may be made only if the redaction is reasonably necessary to protect any of the following:

(I) Public safety.

(II) Trade secrets, as defined in Section 3426.1 of the Civil Code.

(III) Confidential information pursuant to state and federal law.

(B) The developer shall grant to the Attorney General access to the unredacted auditor's report upon request.

(C) An auditor's report disclosed to the Attorney General pursuant to this paragraph is exempt from the California Public Records Act (Division 10 (commencing with Section 7920.000) of Title 1 of the Government Code).

(7) An auditor shall not knowingly make a material misrepresentation in the auditor's report.

(f) (1) (A) A developer of a covered model shall annually submit to the Attorney General a statement of compliance with the requirements of this section signed by the chief technology officer, or a more senior corporate officer, that meets the requirements of paragraph (2).

(B) This paragraph applies if the covered model or any covered model derivatives controlled by the developer remain in commercial or public use or remain available for commercial or public use.

(2) In a statement submitted pursuant to paragraph (1), a developer shall specify or provide, at a minimum, all of the following:

(A) An assessment of the nature and magnitude of critical harms that the covered model or covered model derivatives may reasonably cause or materially enable and the outcome of the assessment required by paragraph (1) of subdivision (b).

(B) An assessment of the risk that compliance with the safety and security protocol may be insufficient to prevent the covered model or covered model derivatives from causing or materially enabling critical harms.

(C) A description of the process used by the signing officer to verify compliance with the requirements of this section, including a description of the materials reviewed by the signing officer, a description of testing or other evaluation performed to support the statement and the contact information of any third parties relied upon to validate compliance.

(g) A developer of a covered model shall report each artificial intelligence safety incident affecting the covered model, or any covered model derivatives controlled by the developer, to the Attorney General within 72 hours of the developer learning of the artificial intelligence safety incident or within 72 hours of the developer learning

facts sufficient to establish a reasonable belief that an artificial intelligence safety incident has occurred.

(h) (1) A developer shall submit to the Attorney General a statement described by subdivision (f) no more than 30 days after using a covered model or covered model derivative for a purpose not exclusively related to the training or reasonable evaluation of the covered model or compliance with state or federal law or making a covered model or covered model derivative available for commercial or public, or foreseeably public, use for the first time.

(2) This subdivision does not apply with respect to a covered model derivative if the developer submitted a statement described by subdivision (f) for the applicable covered model from which the covered model derivative is derived.

(i) In fulfilling its obligations under this chapter, a developer shall consider industry best practices and applicable guidance from the U.S. Artificial Intelligence Safety Institute, National Institute of Standards and Technology, the Government Operations Agency, and other reputable standard-setting organizations.

(j) (1) This section shall not apply to products or services to the extent that the requirements would strictly conflict with the terms of a contract with a federal government entity and a developer of a covered model.

(2) This section applies to the development, use, or commercial or public release of a covered model or covered model derivative for any use that is not the subject of a contract with a federal government entity, even if that covered model or covered model derivative has already been developed, trained, or used by a federal government entity.

22604. (a) A person that operates a computing cluster shall implement written policies and procedures to do all of the following when a customer utilizes compute resources that would be sufficient to train a covered model:

(1) Obtain the prospective customer’s basic identifying information and business purpose for utilizing the computing cluster, including all of the following:

(A) The identity of the prospective customer.

(B) The means and source of payment, including any associated financial institution, credit card number, account number, customer identifier, transaction identifiers, or virtual currency wallet or wallet address identifier.

(C) The email address and telephonic contact information used to verify the prospective customer’s identity.

(2) Assess whether the prospective customer intends to utilize the computing cluster to train a covered model.

(3) If a customer repeatedly utilizes computer resources that would be sufficient to train a covered model, validate the information initially collected pursuant to paragraph (1) and conduct the assessment required pursuant to paragraph (2) prior to each utilization.

(4) Retain a customer’s Internet Protocol addresses used for access or administration and the date and time of each access or administrative action.

(5) Maintain for seven years and provide to the Attorney General, upon request, appropriate records of actions taken under this section, including policies and procedures put into effect.

(6) Implement the capability to promptly enact a full shutdown of any resources being used to train or operate models under the customer’s control.

(b) A person that operates a computing cluster shall consider industry best practices and applicable guidance from the U.S. Artificial Intelligence Safety Institute, National Institute of Standards and Technology, and other reputable standard-setting organizations.

(c) In complying with the requirements of this section, a person that operates a computing cluster may impose reasonable requirements on customers to prevent the collection or retention of personal information that the person that operates a computing cluster would not otherwise collect or retain, including a requirement that a corporate customer submit corporate contact information rather than information that would identify a specific individual.

22606. (a) The Attorney General may bring a civil action for a violation of this chapter and to recover all of the

following:

(1) For a violation that causes death or bodily harm to another human, harm to property, theft or misappropriation of property, or that constitutes an imminent risk or threat to public safety that occurs on or after January 1, 2026, a civil penalty in an amount not exceeding 10 percent of the cost of the quantity of computing power used to train the covered model to be calculated using average market prices of cloud compute at the time of training for a first violation and in an amount not exceeding 30 percent of that value for any subsequent violation.

(2) For a violation of Section 22607 that would constitute a violation of the Labor Code, a civil penalty specified in subdivision (f) of Section 1102.5 of the Labor Code.

(3) For a person that operates a computing cluster for a violation of Section 22604, for an auditor for a violation of paragraph (6) of subdivision (e) of Section 22603, or for an auditor who intentionally or with reckless disregard violates a provision of subdivision (e) of Section 22603 other than paragraph (6) or regulations issued by the Government Operations Agency pursuant to Section 11547.6 of the Government Code, a civil penalty in an amount not exceeding fifty thousand dollars ($50,000) for a first violation of Section 22604, not exceeding one hundred thousand dollars ($100,000) for any subsequent violation, and not exceeding ten million dollars ($10,000,000) in the aggregate for related violations.

(4) Injunctive or declaratory relief.

(5) (A) Monetary damages.

(B) Punitive damages pursuant to subdivision (a) of Section 3294 of the Civil Code.

(6) Attorney’s fees and costs.

(7) Any other relief that the court deems appropriate.

(b) In determining whether the developer exercised reasonable care as required in Section 22603, all of the following considerations are relevant but not conclusive:

(1) The quality of a developer’s safety and security protocol.

(2) The extent to which the developer faithfully implemented and followed its safety and security protocol.

(3) Whether, in quality and implementation, the developer’s safety and security protocol was inferior, comparable, or superior to those of developers of comparably powerful models.

(4) The quality and rigor of the developer’s investigation, documentation, evaluation, and management of risks of critical harm posed by its model.

(c) (1) A provision within a contract or agreement that seeks to waive, preclude, or burden the enforcement of a liability arising from a violation of this chapter, or to shift that liability to any person or entity in exchange for their use or access of, or right to use or access, a developer’s products or services, including by means of a contract of adhesion, is void as a matter of public policy.

(2) A court shall disregard corporate formalities and impose joint and several liability on affiliated entities for purposes of effectuating the intent of this section to the maximum extent allowed by law if the court concludes that both of the following are true:

(A) The affiliated entities, in the development of the corporate structure among the affiliated entities, took steps to purposely and unreasonably limit or avoid liability.

(B) As the result of the steps described in subparagraph (A), the corporate structure of the developer or affiliated entities would frustrate recovery of penalties, damages, or injunctive relief under this section.

(d) Penalties collected pursuant to this section by the Attorney General shall be deposited into the Public Rights Law Enforcement Special Fund established pursuant to Section 12530 of the Government Code.

(e) This section does not limit the application of other laws.

22607. (a) A developer of a covered model or a contractor or subcontractor of the developer shall not do any of the following:

(1) Prevent an employee from disclosing information to the Attorney General or the Labor Commissioner, including through terms and conditions of employment or seeking to enforce terms and conditions of employment if the employee has reasonable cause to believe the information indicates either of the following:

(A) The developer is out of compliance with the requirements of Section 22603.

(B) An artificial intelligence model, including a model that is not a covered model or a covered model derivative, poses an unreasonable risk of causing or materially enabling critical harm, even if the employer is not out of compliance with any law.

(2) Retaliate against an employee for disclosing information to the Attorney General or the Labor Commissioner pursuant to paragraph (1).

(3) Make false or materially misleading statements related to its safety and security protocol in a manner that violates Part 2 (commencing with Section 16600) of Division 7 or any other provision of state law.

(b) An employee harmed by a violation of this subdivision may petition a court for appropriate temporary or preliminary injunctive relief as provided in Sections 1102.61 and 1102.62 of the Labor Code.

(c) (1) The Attorney General or Labor Commissioner may publicly release or provide to the Governor any complaint, or a summary of that complaint, pursuant to this section if the Attorney General or the Labor Commissioner concludes that doing so will serve the public interest.

(2) If the Attorney General or the Labor Commissioner publicly releases a complaint, or a summary of a complaint, pursuant to paragraph (1), the Attorney General or the Labor Commissioner shall redact from the complaint any information that is confidential or otherwise exempt from public disclosure pursuant to the California Public Records Act (Division 10 (commencing with Section 7920.000) of Title 1 of the Government Code) and any information that the Attorney General or the Labor Commissioner determines would likely pose an unreasonable risk to public safety if it were disclosed to the public.

(d) A developer shall provide a clear notice to all employees working on covered models and covered model derivatives of their rights and responsibilities under this section, including the right of employees of contractors and subcontractors to use the developer's internal process for making protected disclosures pursuant to subdivision (e). A developer is presumed to be in compliance with the requirements of this subdivision if the developer does either of the following:

(1) At all times post and display within all workplaces maintained by the developer a notice to all employees of their rights and responsibilities under this section, ensure that all new employees receive equivalent notice, and ensure that employees who work remotely periodically receive an equivalent notice.

(2) No less frequently than once every year, provides written notice to all employees of their rights and responsibilities under this chapter and ensures that the notice is received and acknowledged by all of those employees.

(e) (1) (A) A developer shall provide a reasonable internal process through which an employee may anonymously disclose information to the developer if the employee believes in good faith that the information indicates that the developer has violated any provision of Section 22603 or any other law, or has made false or materially misleading statements related to its safety and security protocol, or failed to disclose known risks to employees, including, at a minimum, a monthly update to the person who made the disclosure regarding the status of the developer's investigation of the disclosure and the actions taken by the developer in response to the disclosure.

(B) The process required by this paragraph shall apply to employees of the developer's contractors and subcontractors working on covered models and covered model derivatives and allow those employees to disclose the same information to the developer that an employee of the developer may disclose and provide the same anonymity and protections against retaliation to the employees of the contractor or subcontractor that apply to disclosures by employees of the developer.

(2) The disclosures and responses of the process required by this subdivision shall be maintained for a minimum of seven years from the date when the disclosure or response is created. Each disclosure and response shall be shared with officers and directors of the developer whose acts or omissions are not implicated by the disclosure or response no less frequently than once per quarter. In the case of a report or disclosure regarding alleged misconduct by a contractor or subcontractor, the developer shall notify the officers

and directors of the contractor or subcontractor whose acts or omissions are not implicated by the disclosure or response about the status of their investigation no less frequently than once per quarter.

(f) This section does not limit protections provided to employees by Section 1102.5 of the Labor Code, Section 12964.5 of the Government Code, or other law.

(g) As used in this section:

(1) "Employee" has the same meaning as defined in Section 1132.4 of the Labor Code and includes both of the following:

(A) Contractors or subcontractors and unpaid advisors involved with assessing, managing, or addressing the risk of critical harm from covered models and covered model derivatives.

(B) Corporate officers.

(2) "Contractor or subcontractor" has the same meaning as in Section 1777.1 of the Labor Code.

22608. The duties and obligations imposed by this chapter 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.

22609. This chapter does not apply to the extent that it is preempted by federal law.

SEC. 4. Section 11547.6 is added to the Government Code, to read:

11547.6. (a) As used in this section, "critical harm" has the same meaning as defined in Section 22602 of the Business and Professions Code.

(b) There is hereby established the Board of Frontier Models. The board shall be housed in the Government Operations Agency and shall be independent of the Department of Technology. The Governor may appoint an executive officer of the board, subject to Senate confirmation, who shall hold the office at the pleasure of the Governor. The executive officer shall be the administrative head of the board and shall exercise all duties and functions necessary to ensure that the responsibilities of the board are successfully discharged.

(c) (1) Commencing January 1, 2026, the Board of Frontier Models shall be composed of nine members, as follows:

(A) A member of the open-source community appointed by the Governor and subject to Senate confirmation.

(B) A member of the artificial intelligence industry appointed by the Governor and subject to Senate confirmation.

(C) An expert in chemical, biological, radiological, or nuclear weapons appointed by the Governor and subject to Senate confirmation.

(D) An expert in artificial intelligence safety appointed by the Governor and subject to Senate confirmation.

(E) An expert in cybersecurity of critical infrastructure appointed by the Governor and subject to Senate confirmation.

(F) Two members who are academics with expertise in artificial intelligence appointed by the Speaker of the Assembly.

(G) Two members appointed by the Senate Rules Committee.

(2) A member of the Board of Frontier Models shall meet all of the following criteria:

(A) A member shall be free of direct and indirect external influence and shall not seek or take instructions from another.

(B) A member shall not take an action or engage in an occupation, whether gainful or not, that is incompatible with the member's duties.

(C) A member shall not, either at the time of the member's appointment or during the member's term, have a financial interest in an entity that is subject to regulation by the board.

(3) A member of the board shall serve at the pleasure of the member's appointing authority but shall serve for no longer than eight consecutive years.

(d) (1) On or before January 1, 2027, and annually thereafter, the Government Operations Agency shall issue regulations to update both of the following thresholds in the definition of a "covered model" to ensure that it accurately reflects technological developments, scientific literature, and widely accepted national and international standards and applies to artificial intelligence models that pose significant risk of causing or materially enabling critical harms.

(2) The updated definition shall contain both of the following:

(A) The initial compute threshold that an artificial intelligence model shall exceed to be considered a covered model.

(B) The fine-tuning compute threshold that an artificial intelligence model shall meet to be considered a covered model.

(3) In developing regulations pursuant to this subdivision, the Government Operations Agency shall take into account all of the following:

(A) The quantity of computing power used to train covered models that have been identified as being reasonably likely to cause or materially enable a critical harm.

(B) Similar thresholds used in federal law, guidance, or regulations for the management of artificial intelligence models with reasonable risks of causing or enabling critical harms.

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

(e) (1) On or before January 1, 2027, and annually thereafter, the Government Operations Agency shall issue regulations to establish binding auditing requirements applicable to audits conducted pursuant to subdivision (e) of Section 22603 of the Business and Professions Code to ensure the integrity, independence, efficiency, and effectiveness of the auditing process. In developing regulations pursuant to this subdivision, the Government Operations Agency shall take into account both of the following:

(A) Relevant standards or requirements imposed under federal or state law or through self-regulatory or standards-setting bodies.

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

(2) Any regulations issued pursuant to paragraph (1) shall, at a minimum, be consistent with guidance issued by the U.S. Artificial Intelligence Safety Institute and the National Institute of Standards and Technology.

(f) (1) On or before January 1, 2027, and annually thereafter, the Government Operations Agency shall issue guidance for preventing unreasonable risks of covered models and covered model derivatives causing or materially enabling critical harms, including, but not limited to, more specific components of, or requirements under, the duties required under Section 22603 of the Business and Professions Code.

(2) Any guidance issued pursuant to paragraph (1) shall, at a minimum, be consistent with guidance issued by the U.S. Artificial Intelligence Safety Institute and the National Institute of Standards and Technology.

(g) Regulations and guidance adopted pursuant to this section shall be approved by the Board of Frontier Models before taking effect.

SEC. 5. Section 11547.6.1 is added to the Government Code, to read:

11547.6.1. (a) There is hereby established in 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 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 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, 2026, 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 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 on potential funding sources.

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

(D) Recommendations on 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) (1) The consortium shall, consistent with state constitutional law, consist of 14 members selected from among all of the following:

(A) Representatives of the University of California and other public and private academic research institutions and national laboratories.

(B) Representatives of impacted workforce labor organizations.

(C) Representatives of stakeholder groups with relevant expertise and experience, including, but not limited to, ethicists, consumer rights advocates, and other public interest advocates.

(D) Experts in technology and artificial intelligence to provide technical assistance.

(E) Personnel from other relevant departments and agencies as necessary.

(2) Eight members of the consortium shall be selected by the Secretary of Government Operations, and the President Pro Tempore of the Senate and the Speaker of the Assembly shall each select three members.

(h) If CalCompute is established within the University of California pursuant to subdivision (c), the University of

California may receive private donations for the purposes of implementing CalCompute.

(i) This section shall become operative only upon an appropriation in a budget act for the purposes of this section.

SEC. 6. 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.

SEC. 7. This act shall be liberally construed to effectuate its purposes.

SEC. 8. The Legislature finds and declares that Section 3 of this act, which adds Chapter 22.6 (commencing with Section 22602) 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 unredacted safety and security protocols and auditor's reports may contain corporate proprietary information or information about covered models and covered model derivatives that could threaten public safety if disclosed to the public.