SB 1047 Section-By-Section Summary

We'll summarize **Definitions, Safety Requirements, Know-Your-Customer Rules, Authorities of the Attorney General, Whistleblower Protections, Government Operations Agency,** and **CalCompute**.

Section 22602. **Definitions.** Selected definitions:

- **"Covered model"** is an AI model trained using greater than 10^26 operations *AND* using compute that would cost at least $100 million when calculated using the average market price of cloud compute. If a covered model is fine-tuned with at least 3×10^25 additional operations at a cost of more than $10 million it is considered to be a *new covered model*. After 2027, the Government Operations Agency may update the computational thresholds. The Government Operations Agency cannot change the monetary thresholds.
- **"Covered model derivative"** Means a copy of or modified version of a covered model.
- **"Critical harm"** means a harm caused or enabled by a covered model or covered model derivative:
    - The creation or use of a chemical, biological, radiological, or nuclear weapon in a manner that results in mass casualties.
    - At least $500 million of damage through a model providing precise instructions for conducting a cyberattack or series of cyberattacks *on critical infrastructure*.
    - At least $500 million of damage by an artificial intelligence model that autonomously engages in conduct that would be criminal if undertaken by a human.
    - Other comparably severe threats to public safety and security.
    - Critical harm does **not** include harms that are caused by the model providing information that is reasonably publicly accessible to an ordinary person (i.e., only includes marginal risks, not bioweapon cookbooks that are easily googleable). It does **not** include harms that aren't caused or *materially* enabled by the developer's creation, storage, use, or release of the covered model. In practice, the latter often means that judges will look at whether the developer's act was a "but-for cause" of the harm: could the malicious actor have just caused the harm anyway, or would they not have done so if not for the developer's activities?
- **"Developer"** is the person that performs the original training of a covered model, whether from scratch or by fine-tuning an existing covered model with a sufficient amount of compute. A person that fine-tunes a model with less than $10M worth of compute is not a developer and has no obligations under the bill.
- **"Full shutdown"** means the cessation of operation of the training of a covered model and all covered models and covered model derivatives *controlled by the developer*. If a developer has released the weights of a model, copies of that model used by other people

are not controlled by the developer, so there is no full shutdown requirement for open weight models outside the developer's control.

Section 22603. **Safety requirements.**

22603(a) **Prior to training:** Before training a covered model, developers must do all of the following prior to and during training:

- **Cybersecurity:** Implement cybersecurity controls to prevent model theft and misuse.
- **Full shutdown:** Implement the capability to perform a full shutdown.
- **Safety and security protocol:** Develop a written safety and security protocol that describes how the developer will prevent unreasonable risk of critical harm and how the covered model will be tested for its ability to do so. The tests must incorporate the possibility that the covered model could be modified to be more dangerous. The protocol must be implemented as written and senior personnel must be designated to ensure compliance.
- **Publishing protocol:** The developer must publish the safety and security protocol with redactions and send it to the Attorney General. The Attorney General can access the unredacted version upon request.

22603(b) **Prior to deployment:** Prior to using or releasing a covered model, the developer must:

- **Assessment:** Assess whether the covered model is reasonably capable of causing or enabling a critical harm and keep track of the details of how the developer tested the model.
- **Safeguards:** Take *reasonable care* to implement appropriate safeguards to prevent the covered model and covered model derivatives from causing or materially enabling a critical harm. "Reasonable care" is an extremely common legal standard that means the care that a reasonable person would have taken in a similar situation. Existing tort law already requires that everyone take reasonable care in their actions, or else they may be found negligent and required to pay damages for resulting foreseeable harms.
- **Attribution:** Take reasonable care to create methods to attribute model actions and resulting harms to the model.
- **Other measures:** Implement any other reasonably necessary measures to prevent unreasonable risk of critical harm.

22603(c) **No unreasonable risk:** A developer must not use or release a covered model or covered model derivative if there is an unreasonable risk of critical harm. "Unreasonable risk" is an extremely common legal standard that means a risk that a reasonable person would not find reasonable.

22603(d) **Reevaluation:** Developers must periodically reevaluate all of the safeguards above.

22603(e) **Auditing:** Developers must retain a third party auditor that produces a report assessing the developer's steps to comply and identifying any instances of noncompliance. The audit report

must be published and transmitted to the Attorney General, with redactions. The Attorney General can access the unredacted version upon request.

22603(f) **Annual statement of compliance:** Developers must annually submit a statement of compliance to the Attorney General that includes information on any steps and evaluations required above.

22603(g) **Incident reporting:** Developers must report AI safety incidents within 72 hours of learning of them.

22603(h) **Statement of compliance for model:** A developer must submit a statement of compliance to the Attorney General within 30 days of using or releasing a covered model.

22603(i) **Consider guidance:** Developers should consider guidance from the Government Operations Agency, National Institute of Standards and Technology, and other reputable standard-setting organizations.

Section 22604. **Know-your-customer rules.**

Providers of computing resources must implement know-your-customer rules for all customers using enough computing resources to train a covered model and assess whether customers intend to deploy a covered model. The developer must provide records of actions taken under this section to the Attorney General upon request. The developer does *not* need to provide the customer records themselves to the Attorney General.

Section 22606. **Authorities of the Attorney General.**

If the Attorney General finds that a person is violating the act, they may bring a civil action. For auditors who have knowingly violated the requirements around auditing or lied on their audit reports, or developers who violate the whistleblower protections, or compute providers who violate know-your-customer rules, courts can order $50,000 fines per violation. For all other violations, fines can only be ordered for violations that cause harm or pose an imminent risk or threat to public safety. Courts can also order injunctions (an order for a developer to modify or cease its behavior). Finally, courts can order civil damages to redress harms caused by violations.

Section 22607. **Whistleblower protections.**

Establishes whistleblower protections for employees of frontier AI developers to report violations of this act. Employees can also report risky situations that do not involve any violation of the law.

Section 22608. **Cumulative duties.**

The duties under this law are cumulative with duties imposed by other laws.

Section 4. **Government Operations Agency**

- Empowers the existing California Government Operations Agency with three duties:
    - Changing compute thresholds. The monetary thresholds cannot be changed.
    - Issuing binding auditing regulations to ensure the integrity, independence, efficiency, and effectiveness of the audit process.
    - Issuing voluntary safety guidance for developers.
- **The Board of Frontier Models:** Created to oversee the work of the Government Operations Agency. It is made up of experts appointed by the Governor and the legislature, and it must approve all regulations and guidance.
- **Members of the Board**: Five appointed by the Governor and confirmed by the Senate: a member from the open source community; member of the AI industry; an expert in chemical, biological, radiological or nuclear weapons; an expert in AI safety; an expert in the cybersecurity of critical infrastructure. Two members each appointed by the Senate Rules Committee and the Speaker of the Assembly.

Section 5. **CalCompute.**

- Authority to create a consortium to develop a framework for the creation of CalCompute, a public cloud computing resource.
- The consortium shall make efforts to ensure CalCompute should be hosted in the University of California.
- The Consortium must submit a report to the legislature by 2026 for the creation of CalCompute.

Important points:

- There are **no criminal penalties** in SB 1047 (previously, the only criminal penalty was for intentionally lying on a business statement, but that penalty has been removed).
- There are **no obligations** for anyone who is not a developer or a person who operates a computing cluster, except that contractors and subcontractors of developers can't retaliate against whistleblowers.
- Developers are free to determine how they will implement their safety and security protocols in order to prevent critical harms. SB 1047 does not prescribe specific ways to, for example, attain adequate cybersecurity or perform full shutdowns.
- Developers **already have liability** for failing to take reasonable care to prevent harm in ordinary tort law. SB 1047 provides more clarity than existing tort law.