Comparison of California's SB 53 and New York's RAISE Act: Foundation Model Frameworks

Created by: Justine Gluck, AI Policy Analyst

Overview: On September 29, Governor Newsom signed **SB 53**, or the "**Transparency in Frontier Artificial Intelligence Act (TFAIA),"** introduced by Sen. Weiner (D). The bill is California's second major legislative effort to address the safety and oversight of AI frontier models, following **SB 1047**, a prior version of TFAIA that was vetoed by Gov. Newsom in 2024. The bill shares many similarities with **New York's A 6953**, or the **"Responsible AI Safety and Education (RAISE) Act,"** passed by the New York Legislature on June 12 and also awaiting Governor signature. While both RAISE and TFAIA seek to regulate large frontier model developers by imposing disclosure and transparency requirements, including the adoption of written protocols/ frameworks and the reporting of safety incidents, the two bills diverge in several key respects that may affect entities seeking to comply with the laws:

1. **Strict Liability:** The RAISE Act includes a prohibition on the deployment of frontier models that pose an "unreasonable risk of critical harm," a provision not included in TFAIA.
2. **Scope:** TFAIA uses wider definitions for "catastrophic risk" and separately defines "frontier developers" and "large developers," which may subject more entities to compliance than the RAISE Act.
3. **Requirements:** TFAIA imposes more prescriptive requirements and creates additional obligations such as whistleblower protections and transparency reports.
4. **Enforcement:** TFAIA sets a penalty system with lower penalties compared to the RAISE Act and grants the California Department of Technology broader regulatory authority.

This comparative analysis considers these similarities and differences between California's and New York's frameworks, covering (1) scope; (2) compliance requirements; (3) other requirements; and (4) enforcement.

Red text indicates key differences between the TFAIA and the RAISE Act.

	California SB 53 (TFAIA)	New York A 6953 (RAISE Act)	Comparison
	Scope		
Scope	**Foundation Model** means an AI model that is: 1. Trained on a broad data set; 2. Designed for generality of output; and 3. Adaptable to a wide range of distinctive tasks (Sec. 22757.11 (f)). **Frontier Model** means a foundation model that was trained using a quantity of computing power greater than 10^26 computational operations (e.g., integer or floating-point operations) (Sec. 22757.11 (i)). - This quantity of computing power includes the original training run and subsequent modifications to the foundation model. **Frontier Developer:** A person who has trained or initiated the training of a frontier model which the person intends to use at least as 10^26 computing power (Sec. 22757.11	**Frontier Model** is *either* of the following: a. An artificial intelligence model trained using greater than 10^26 computational operations (e.g., integer or floating-point operations), the compute cost of which exceeds one hundred million dollars; or b. (b) an artificial intelligence model produced by applying knowledge distillation to a frontier model, provided that the compute cost exceeds five million dollars (Sec. 1420(6)). Frontier models must also be developed, deployed, or operated in whole or in part in New York State (Sec. 1424). **Large Developer:** A person that has trained at least one frontier model and has spent over $100 million in compute costs in aggregate in training frontier models:	***RAISE and TFAIA define similar scopes but have different approaches to monetary thresholds, which may offer different scopes of applicability.*** **Frontier Model:** Both bills use the same compute benchmark, targeting the most advanced and resource-intensive systems. But TFAIA broadly includes cumulative compute used not only in initial training but *also* in any fine-tuning or modifications. RAISE adds a distillation clause, extending coverage to smaller models derived from large ones, but pairs this with a high compute cost threshold that narrows developer scope. **Cost Threshold:** The RAISE Act defines "large developers" as those spending $100M+ on compute, targeting companies training at very high

	(h)). **Large Frontier Developer:** A frontier developer that together with its affiliates collectively had gross revenues in excess of $500 million (Sec. 22757.11 (j)).	- Accredited colleges and universities are not included within scope to the extent they engage in academic research (Sec. 1420(9)).	compute levels. By contrast, TFAIA uses a $500M+ annual revenue threshold, which could bring a distinct group of companies into scope compared to the RAISE Act. Neither is scoped to apply to small developers. **Developer Categories:** TFAIA distinguishes between "frontier developers" and "large frontier developers," applying different requirements to each group. The cost threshold applies only to the latter, meaning some obligations extend to developers regardless of cost. RAISE, by contrast, scopes requirements solely to "large developers." **Extraterritorial Application:** TFAIA doesn't explicitly restrict scope to California-based developers and was framed by Newsom as a blueprint beyond state borders. However, RAISE, applies only to models "developed, deployed, or operated in whole or in part in New York." **Academia Exemption:** RAISE explicitly exempts universities; TFAIA provides no similar exemption.
Key Terms	**Catastrophic Risk:** Foreseeable and material risk that a frontier developer's development, storage, use, or deployment of a frontier model will materially contribute to the death or serious injury of 50+ or at least $1 billion of damages to loss of property (tangible and intangible) arising from a single incident involving a frontier model doing any of the following: a. Provide expert-level assistance in the creation or release of a chemical, biological, radiological, or nuclear weapon; b. Engage in conduct with no meaningful human intervention that would, if committed by a human, constitute a crime or is a cyberattack; or c. Evade the control of its frontier developer or user (Sec. 22757.11 (c)(1)). The loss of value of equity does not count as damage to or loss of property (Sec. 1107.2).	**Critical Harm:** Death or serious injury of 100+ people or at least $1 billion of damages to rights in money or loss of property caused or materially enabled by a large developer's creation, use, storage, or release of a frontier model, through either: a. Creation or use of a chemical, biological, radiological, or nuclear weapon; or b. An AI model engaging in conduct that does both: i. Acts with limited human intervention ii. Would, if committed by a human, constitute a crime (Sec. 1420(7)). A harm inflicted by intervening human actor shall not be deemed to result from a developer's activities unless such activities were a substantial factor in bringing about the harm, the intervening human actor's conduct was reasonably foreseeable as a probable consequence of the developer's activities, and could have been reasonably	***TFAIA defines risk more broadly and includes fewer liability limitations.*** **Scope of Risk and Harm:** Both bills set high thresholds for risk, but TFAIA's threshold is higher (e.g. 50+ deaths vs. 100+ deaths). TFAIA's "catastrophic risk" also includes additional model behaviors, such as evading developer control. **Dangerous Capabilities:** TFAIA includes any model providing "expert-level assistance" in creating or using a weapon, a lower bar than RAISE's requirement that the model "cause or materially enable" the harm. **Liability Limitations:** The RAISE Act contains heightened thresholds for liability, requiring that harm be a "probable consequence" of the

	Catastrophic risk does not include a foreseeable and material risk from any of the following: a. Information that a frontier model outputs if it is publicly accessible in a substantially similar form from another source; b. Lawful activity of the federal government; and c. Harm caused by a frontier model in combination with other software if the frontier model did not materially contribute (Sec. 22757.11 (c)(2)). **Deploy:** to make a frontier model available to third-party for use, modification, and copying (except developing/evaluating frontier model) (Sec. 22757.11 (e)).	prevented or mitigated (Sec. 1420(7)). **Deploy:** to use a frontier model or make it available to third-party for use, modification, and copying (except training/ developing/evaluating the frontier model, or complying with federal or state laws) (Sec. 1420(5)).	developer's activities, that the developer's actions be a "substantial factor," and that harm couldn't have been "reasonably prevented." TFAIA lacks these limitations and uses a broader standard.
Compliance Requirements			
Frontier AI Framework/ Safety and Security Protocol	**Content:** **Frontier AI framework** means documented technical and organizational protocols to manage, assess, and mitigate catastrophic risks (Sec. 22757.11 (g)). **A large frontier developer shall implement, comply with, and clearly and conspicuously publish a frontier AI framework** that describes how the large frontier developer approaches the following: 1. Incorporating national standards, international standards, and industry-consensus best practices into its frontier AI framework; 2. Defining and assessing thresholds used to assess whether frontier model has capabilities that could pose a catastrophic risk; 3. Applying mitigations to address potential for catastrophic risks; 4. Reviewing assessments and adequacy of mitigations as part of decision to deploy the frontier model; 5. Using third parties to assess the potential for catastrophic risks and effectiveness of mitigations; 6. Updating frontier AI framework, including criteria triggering updates and how the developer	**Content:** **Safety and security protocol must:** a. Describe reasonable protections and procedures that would appropriately reduce the risk of critical harms; b. Describe reasonable cybersecurity protections for frontier models that, if successfully implemented, appropriately reduce the risk, unauthorized access to, or misuse of, the frontier models; c. Describe in detail the testing procedure to evaluate if the frontier model poses an unreasonable risk of critical harm; d. Enable developer or third party to comply with article's requirements; and e. Designate senior personnel for compliance (Sec. 1420(12)). **Administration:** **Before deploying frontier model, a large developer shall:** a. Implement a ***written safety and security protocol***; b. Retain an unredacted copy of the safety and security protocol for as long as the frontier model is deploying, plus five years; c. Conspicuously publish a copy of the safety and	***Both bills mandate written frameworks and public disclosure, but TFAIA specifies more content requirements while RAISE layers in stricter reporting obligations, including information on testing.*** **Protocol Content:** TFAIA is more prescriptive, requiring detailed documentation of governance structures, mitigation processes, and alignment with national/international standards. It also explicitly covers catastrophic risk from *internal use* of models, raising the scope of compliance obligations. RAISE includes similar core categories (cybersecurity, safeguards, mitigation) but is less detailed, potentially allowing more discretion but also leaving firms with greater uncertainty about what will satisfy compliance. **Compliance Personnel:** RAISE requires designation of senior personnel for protocol compliance, ensuring clear accountability. TFAIA has no such requirement, possibly offering developers more discretion in assigning roles.

	determines when its frontier models are substantially modified enough to require disclosures; 7. Cybersecurity practices and how they secure unreleased model weights; 8. Identifying and responding to critical safety incidents; 9. Instituting internal governance practices for implementation of these processes; and 10. Assessing and managing catastrophic risk from internal use (Sec. 22757.12 (a)). **Administration**: Large frontier developer shall review and, if appropriate, update its frontier AI framework at least annually (Sec. 22757.12 (b)(1)). If a large frontier developer makes a material modification to its frontier AI framework, they must clearly and conspicuously publish the framework and justification within 30 days (Sec. 22757.12 (b)(2)). **Redactions:** Frontier developers may make redactions to the framework *(and transparency report)* to protect trade secrets, cybersecurity, national security. To the extent permitted, must publicly describe redactions and retain unredacted information for 5 years (Sec. 22757.12 (f)).	security protocol (with appropriate redactions) and transmit this document to the attorney general and division of Homeland Security and Emergency Services (DHSES) d. Record, and when reasonably possible, retain, information on the tests and test results used in any assessment of the frontier model; and e. Implement appropriate safeguards to prevent unreasonable risk of critical harm (Sec. 1421(1)). A large developer shall conduct an annual review of protocols to account for any changes in the capabilities of their frontier models and industry best practices and, if necessary, make protocol modifications (Sec. 1421(3)).	**Testing Requirements:** RAISE requires transparency around testing procedures in the safety protocol, but offers little detail on what tests are required or how they should be conducted. TFAIA drops direct testing language, referring instead to "assessments." Neither bill explicitly mandates that specific tests be performed, leaving open questions about what level of testing is actually necessary for compliance. **Timing & Updates:** Both require annual reviews. RAISE goes further by mandating protocols be in place before deployment and transmitted to both the AG and DHSES, creating additional pre-deployment obligations. TFAIA requires the framework to be re-published within 30 days of modifications, which may force companies to account for faster revision cycles than under RAISE.
Transparency Report	Before, or concurrently with, deploying a frontier model a frontier developer shall clearly and conspicuously publish on its website a **transparency report** containing all of the following: a. Website of the frontier developer; b. Mechanism that allows a natural person to communicate with the frontier developer; c. Release date of the frontier model; d. Languages supported by the frontier model; e. Modalities of output supported by frontier model; f. Intended uses of frontier model; g. Restrictions or conditions on uses of the frontier model (Sec. 22757.12 (c)(1)).	**No transparency report requirement.** ***Many of TFAIA's transparency requirements are included within the RAISE Act's safety and security protocol, such as a prohibition for developers to make materially false statements.***	***Only TFAIA requires frontier developers to publish detailed transparency reports.*** **Pre-Deployment Transparency:** TFAIA mandates that frontier developers publish a transparency report **before or concurrently with deployment**, similar to RAISE's pre-deployment safety protocols. **Scope:** Unlike many other TFAIA obligations that apply only to "large frontier developers," the transparency report requirement applies to all frontier developers, meaning even smaller firms that meet the "frontier developer" definition face compliance requirements.

	Before, or concurrently with, deploying a frontier model, a frontier developer shall include in the **transparency report** summaries of all of the following: a. Assessments of catastrophic risks conducted pursuant the frontier AI framework; b. Assessment results; c. Involvement of third-party evaluators; d. Other steps to fulfill requirements of the frontier AI framework (Sec. 22757.12 (c)(2)). Frontier developers can publish this information as part of a larger document, like a system or model card (Sec. 22757.12 (c)(3)). Frontier developers encouraged, but not required, to make disclosures that are consistent or superior to industry best practices (Sec. 22757.12 (c)(4)). Large frontier developers shall transmit to the Office of Emergency Services (OES) a summary of any assessment of catastrophic risk resulting from internal use of its frontier models every three months (Sec. 22757.12 (d)).		**Internal Use:** TFAIA requires publication of a summary of any catastrophic risk assessment stemming from **internal use** of a foundation model, broadening the scope of required transparency. The RAISE Act does not specify "internal use" requirements. **Integration with Practices:** Firms can incorporate TFAIA disclosures into existing documents like system or model cards, which may help ease compliance. TFAIA also encourages alignment with industry best practices, though the lack of clear benchmarks may create compliance uncertainty.
Prohibition on Deployment	**N/A**	**A large developer shall not deploy a frontier model if doing so would create an unreasonable risk of critical harm** (Sec. 1421(2)).	***The RAISE Act contains a strict prohibition against deployment of models with critical risk.*** **Prohibition on Deployment:** The RAISE Act bars deployment if there's an "unreasonable risk of critical harm," with carveouts for training, evaluation, or legal compliance encompassed in the bill's definition of "deploy." However, TFAIA does not include this prohibition, focusing instead on transparency and reporting requirements. *SB 1047 originally included this prohibition.*
Disclosure of Safety Incidents	The OES will establish a mechanism for the frontier developer or member of the public to report a critical safety incident that includes: 1. The date of the safety incident; 2. Reasons the incident qualifies as a safety incident;	A large developer shall disclose each safety incident affecting the frontier model to the AG and DHSES within 72 hours of the large developer learning of the safety incident or facts sufficient to establish a reasonable belief that a safety incident has occurred (Sec. 1421(4)).	***TFAIA allows public reporting of safety incidents and offers developers more time to disclose non-imminent risks, while RAISE imposes a shorter 72-hour window and uses qualifiers that raise the bar for a reportable incident.***

	3. A short and plain statement describing the safety incident; and 4. Whether the incident was associated with internal model use (Sec. 22757.13 (a)). A frontier developer shall report any critical safety incident within 15 days of discovery (Sec. 22757.13 (c)(1)). If a frontier developer discovers a critical safety incident poses an imminent risk of danger of death or serious injury, they shall disclose that incident no later than 24 hours to an authority (Sec. 22757.13 (c)(2)). A frontier developer is encouraged, but not required, to report critical safety incidents pertaining to foundation models that are not frontier models (Sec. 22757.13 (c)(4)). **Critical Safety Incident** means any of the following: 1. Unauthorized access to, modification of the model weights of a foundation model that results in death or bodily injury; 2. Harm resulting from the materialization of a catastrophic risk; 3. Loss of control of a frontier model causing death or bodily injury, or loss of property; or 4. A foundation model that employs deceptive techniques to evade the controls or monitoring of its frontier developer in a manner that demonstrates materially increased risk (Sec. 22757.11(d)). The AG/ OES may transmit reports of safety incidents to the Legislature, Gov., federal government, or agencies. Risks related to trade secrets, public safety, cybersecurity, or national security shall be strongly considered when transmitting reports (Sec. 22757.13 (e)(1)). The OES shall produce an anonymized annual report with information on critical safety incidents and transmit the report to the Legislature and Governor (Sec. 22757.13 (g)).	Such disclosure shall include: a. The date of the safety incident; b. Reasons the incident qualifies as a safety incident; and c. A short and plain statement describing the safety incident (Sec. 1421(4)). **Safety Incident:** A known incident of critical harm or an incident of the following that provides demonstrable evidence of an increased risk of critical harm: a. A frontier model autonomously engaging in behavior other than at the request of a user; b. Theft, misappropriation, malicious use, inadvertent release, unauthorized access, or escape of the model weights of a frontier model; c. Critical failure of any technical or administrative controls; and d. Unauthorized use of frontier model (Sec. 1420(13)).	**Reporting Timeline:** RAISE requires disclosure within 72 hours or "reasonable belief," while TFAIA allows 15 days, unless there is an imminent risk of "danger of death or serious physical injury" and the timeline shortens to 24 hours. **Threshold:** RAISE uses a "reasonable belief" standard and requires "demonstrable evidence" of increased risk, raising the bar for what qualifies as incident reporting and requiring action even in the absence of confirmed harm. **Incident Scope:** Both include comparable incidents involving unauthorized access, misuse, or loss of control. **Public Reporting:** Only TFAIA requires the AG to establish a mechanism for the public to report safety incidents, expanding oversight beyond developers. While this could enhance transparency, it may also raise concerns for developers about unverified public claims.
Other Requirements			

Whistleblower Protections	A frontier developer shall not adopt a policy or contract that retaliates against or prevents a covered employee from disclosing information to the AG, or other authority, if the covered employee has reasonable cause to believe developer's activities pose a specific and substantial danger to the public health or safety resulting from a catastrophic risk or have violated the Act. **Covered Employee:** an employee responsible for assessing, managing, or addressing risk of critical safety incidents (Sec. 1107(b)). Frontier developer shall provide a clear notice to all covered employees of their rights/ responsibilities, by: 1. Displaying at all times within the workplace a notice to all covered employees of their rights; or 2. Annually, providing written notice to covered employees of their rights (Sec. 1107.1 (e)). The frontier developer shall provide a reasonable internal process for an employee to anonymously disclose information and provide monthly update to the discloser. The large developer has the burden of proof. Courts must consider direct harm and potential chilling effect on other employees (Sec. 1107.1 (h)).	**N/A**	***TFAIA establishes whistleblower protections; the RAISE Act does not address whistleblowers.*** TFAIA prohibits retaliation against employees or contractors who report activity from a catastrophic risk, mandates notice of employee rights, and requires anonymous internal reporting channels.
Enforcement			
Enforcement	The AG may bring a civil action against a large frontier developer that fails to publish a document, report an incident, or comply with its own frontier AI framework. Civil penalty up to $1 million per violation, dependent on the severity of the violation (Sec. 22757.15(a)(b)). Before January 1, 2027, the **Department of Technology may make recommendations about updating the definitions of "frontier model," "frontier developer," and "large frontier developer"** to ensure it reflects technological developments, submitting a report to the Legislature (Sec. 22757.15 (a)).	The Attorney General (AG) may bring a civil action for a violation, determined based on the severity of the violation: a. A civil penalty in an amount not exceeding $10 million for a first violation and $30 million for any subsequent violation; b. Injunctive or declaratory relief (Sec. 1422(1)). No private right of action (Sec. 1422(2)).	***TFAIA sets lower penalties compared to the RAISE Act, offers definitional adaptability, and AG reporting requirements.*** **Enforcement Mechanism:** Both bills authorize the AG to bring civil actions for violations. Neither bill includes a private right of action, though RAISE explicitly prohibits one, unlike TFAIA. **Penalties:** RAISE sets significantly higher penalties, up to $10 million for a first violation and $30 million for subsequent ones, based on

	The AG shall produce a report about reports from employees responsible for addressing critical safety incidents and submit the report to the Legislature (Sec. 22757.14 (d)).		severity. TFAIA also considers severity, but with lower penalties capped at $1 million per violation, signaling a more modest enforcement. Neither bill offers any affirmative defense or safe harbor. **Definitional Adjustment:** TFAIA uniquely empowers the Department of Technology to recommend updates to statutory definitions to keep pace with technological change. While these recommendations require legislative adoption, this mechanism builds in definitional adaptability absent in RAISE. Earlier drafts of TFAIA granted the AG direct rulemaking authority to revise definitions, but was narrowed in the final bill.

Key Differences Between California's SB 1047 (2024, vetoed)

1. **Pre-Training Requirements:** SB 1047 would have required developers to implement safety protocols, cybersecurity protections, and full shutdown capabilities before beginning initial training of a covered model.
    a. *Neither RAISE nor TFAIA imposes pre-training obligations; both focus on deployment-stage requirements.*

2. **Full Shutdown:** SB 1047 would have mandated that covered models include a full shutdown capability as a safety mechanism, developed pre-training.
    a. This requirement is not present in either RAISE or TFAIA.

3. **Retention of Tests:** SB 1047 would have required developers to retain testing procedures and results for the full duration of the model, plus five years.
    a. Neither RAISE nor TFAIA includes comparable retention obligations for testing, although RAISE includes a similar requirement for the unredacted safety and security protocol.

4. **Third-Party Audits:** SB 1047 would have required developers to retain an independent third-party auditor annually to assess internal controls and compliance.
    a. *RAISE and TFAIA do not contain any third-party audit requirements.*

5. **72-Hour Safety Incident Reporting:** SB 1047 would have required reporting a safety incident to the Attorney General within 72 hours of forming a "reasonable belief" that it occurred.
    a. RAISE mirrors this standard; TFAIA provides a longer 15-day window and lacks the "reasonable belief" trigger, but limits to 24 hours for imminent risks.

6. **Civil Penalties:** SB 1047 would have set penalties up to 10% of compute cost used to train the model (30% for subsequent violations), scaled to the harm's severity.
    a. RAISE sets flat penalties ($10M/$30M), while TFAIA caps penalties at $1 million per violation, although both use severity of harm as a metric.