October 2, 2025

If you have any questions regarding the matters discussed in this memorandum, please contact the attorneys listed on the last page or call your regular Skadden contact.

One Manhattan West
New York, NY 10001
212.735.3000

Landmark California AI Safety Legislation May Serve as a Model for Other States in the Absence of Federal Standards

Executive Summary

- **What's new:** California has enacted a landmark AI law that establishes the nation's first comprehensive framework for transparency, safety and accountability in the development and deployment of advanced AI models.
- **Why it matters:** The legislation requires developers of "frontier" AI models to publish detailed safety frameworks, report serious safety incidents and strengthen whistleblower protections. It contrasts with the light regulatory approach adopted by the Trump administration, and comes after Congress failed to impose a moratorium on state AI laws.
- **What to do next:** Developers of AI models will need to assess whether their models and revenue meet the law's definitions of "frontier model" and "large frontier developer." If they do, the developer will need to prepare to create a "frontier AI framework" to manage risks, and develop protocols for identifying and reporting "critical safety incidents" to the state. HR policies, employment agreements and internal reporting channels may also need to be revised to comply with the new whistleblower protections.

On September 29, 2025, California Governor Gavin Newsom signed into law Senate Bill 53 (SB 53), known as the Transparency in Frontier Artificial Intelligence Act (TFAIA). This landmark legislation establishes the nation's first comprehensive framework for transparency, safety and accountability in the development and deployment of advanced artificial intelligence (AI) models.

TFAIA establishes a new regulatory floor for developers of "frontier" AI models, which could become a template for other states. The law's core provisions require such developers to publish detailed safety framework, report serious safety incidents and strengthen whistleblower protections for employees who flag "catastrophic" risks or violations of the law.

TFAIA was signed only a few months after a failed attempt in Congress to impose a 10-year moratorium on most state AI laws, and it runs contrary to the light regulatory touch advocated by the Trump administration. See our July 30, 2025, client alert, "White House Releases AI Action Plan: Key Legal and Strategic Takeaways for Industry."

Landmark California AI Safety Legislation May Serve as a Model for Other States in the Absence of Federal Standards

For developers of AI models, this law introduces new legal and operational requirements that demand prompt attention. In the final section of this article, we outline the preparatory steps developers can take before the law takes effect, which is expected to be on January 1, 2026.

Background of the Legislation

TFAIA creates a multi-faceted regulatory scheme for the developers of certain AI frontier models. The law does not create new liability for harms caused by AI systems but focuses instead on transparency and risk management.[1] In signing the bill, Governor Newsom stated that it further establishes California as a leader in "safe, secure and trustworthy artificial intelligence."

In 2024, Governor Newsom vetoed California's first attempt to enact an omnibus AI safety law, finding that the law was overly burdensome to AI developers. Subsequently, the governor convened the Joint California Policy Working Group on AI Frontier Models to provide recommendations in a number of areas, including transparency measures and whistleblower protections.

TFAIA is, in part, a product of that Working Group's recommendations. While other states have enacted specific AI regulations, TFAIA is the first state law to impose broad safety requirements on AI developers. The law instructs the California Department of Technology to recommend updates to the law on an annual basis

Compliance Regime for 'Large Frontier Developers'

The law's most significant obligations apply to entities defined as "large frontier developers," a subset of fronter developers that meet the following model compute and revenue levels:

- **Frontier model:** a foundational model trained using more than 1026 integer or floating-point operations (a measure of computational power).
- **Large frontier developer:** a frontier developer that, along with its affiliates, had annual gross revenues exceeding $500 million in the prior calendar year.

To promote transparency, large frontier developers are required to write, implement and publicly publish a "**frontier AI framework**" on their websites. The framework must describe the company's approach to managing and mitigating catastrophic risks, including its processes for:

- Incorporating national and international standards, as well as industry best practices.
- Assessing whether a model has capabilities that could pose a "catastrophic risk" (defined below).
- Using third parties to assess risks and audit the effectiveness of mitigations.
- Implementing cybersecurity practices to secure unreleased model weights.
- Instituting internal governance practices to ensure compliance with these processes.

"**Catastrophic risk**" is defined as 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 $1 billion in property damage involving a foundation model doing any of the following:

- Providing expert-level assistance in the creation or release of a chemical, biological, radiological or nuclear weapon.
- Engaging in conduct with no meaningful human oversight that is either a cyberattack or, if committed by a human, would constitute the crime of murder, assault, extortion or theft.
- Evading the control of its frontier developer or user.

Noncompliance with publication, reporting or framework requirements can result in civil penalties up to $1 million per violation, enforceable by the California Attorney General.

Safety Incident Reporting to California Authorities

TFAIA establishes a mandatory reporting system for developers and the public to report critical safety incidents to be developed and overseen by the California Office of Emergency Services (OES). "Critical safety incidents" include events such as:

- Unauthorized access to, modification of, or exfiltration of, the model weights of a frontier model that results in death or bodily injury.
- Harm resulting from the materialization of a catastrophic risk.
- The loss of control of a frontier model that causes death or bodily injury.
- A model using deceptive techniques to subvert developer controls that increases catastrophic risk.

Incidents must be reported to OES within 15 days of discovery, or within 24 hours if there is an imminent risk of death or serious injury. These reports are exempt from public records laws in order to protect trade secrets, cybersecurity and public safety. However, the OES will publish annual anonymized and aggregated summaries of incidents.

[1] Even though the TFAIA does not create a new liability regime, developers should be aware that their compliance with the TFAIA "frontier AI framework" publication requirement may still be subject to other general California laws that apply broadly to the AI sector, including California's False Advertising Law, as outlined in the California Office of the Attorney General's Legal Advisory dated January 1, 2025.

Landmark California AI Safety Legislation May Serve as a Model for Other States in the Absence of Federal Standards

Robust Whistleblower Protections

TFAIA creates strong protections for employees who raise safety concerns. Employees responsible for AI risk assessment or management ("covered employees") are protected from retaliation when reporting specific and substantial dangers to public health or safety, or violations of the TFAIA.

Moreover, large frontier developers must provide a "reasonable internal process" that allows for anonymous internal reporting channels and must update whistleblowers monthly on the status of their disclosures. If necessary, employees can seek injunctive relief and attorney's fees for violations, and the burden of proof shifts to the employer once retaliation is alleged.

CalCompute Public Cloud

TFAIA will establish CalCompute, a state-backed public cloud computing cluster designed to foster safe, ethical and equitable AI research and development. A consortium of academic, labor, public interest and technical experts will develop the framework for CalCompute, with a report due to the Legislature by January 1, 2027. CalCompute aims to democratize access to high-performance computing resources, supporting startups, researchers and public interest projects.

What To Do To Prepare

SB 53 creates a new set of compliance obligations. Here are steps developers of advanced AI models should consider taking to prepare:

- **Conduct an applicability analysis.** Companies should determine if they meet the specified thresholds that define "frontier developer" or "large frontier developer."
- **Develop a compliant frontier AI framework.** To the extent applicable, developers should begin drafting the required "Frontier AI Framework" to meet statutory requirements.
- **Establish incident response and reporting protocols.** AI frontier developers should consider how to establish or revise internal procedures to identify, assess and report "critical safety incidents" to the OES within the mandated timelines. Large frontier developers will also want to create the required anonymous internal reporting channel.
- **Update HR policies and employee training:** HR policies, non-disclosure agreements and employment contracts may need to be revised to align with the new whistleblower protections. Companies should make employees aware of their rights under the new law.
- **Monitor regulatory developments:** A number of states, including New York (through the proposed Responsible AI Safety and Education (RAISE) Act), are currently considering similar AI safety laws. California continues to be a state leader in enacting AI-related regulation. AI developers should closely monitor developments at the state level. See our September 27, 2024, client alert "California Enacts New Laws to Combat AI-Generated Deceptive Election Content."

Contacts

Ken D. Kumayama
Partner / Palo Alto
650.470.4553
ken.kumayama@skadden.com

Stuart D. Levi
Partner / New York
212.735.2750
stuart.levi@skadden.com

William E. Ridgway
Partner / Chicago
312.407.0449
william.ridgway@skadden.com

David A. Simon
Partner / Washington, D.C.
202.371.7120
david.simon@skadden.com

Don L. Vieira
Partner / Washington, D.C.
202.371.7250
donald.vieira@skadden.com

Joshua Silverstein
Counsel / Washington, D.C.
202.371.7148
joshua.silverstein@skadden.com

Michael Tian
Law Clerk / Washington, D.C.
202.371.7595
michael.tian@skadden.com