

California Enacts Comprehensive AI Safety Law: What AI Developers Need to Know

OCTOBER 24, 2025

[MATTHEW BERLIN](#), [NATASHA WEIS](#)

Share This Page [EMAIL](#) [LINKEDIN](#) [X](#) [FACEBOOK](#)

California Governor Gavin Newsom signed Senate Bill 53, “The Transparency in Frontier Intelligence Act” (TFAIA), establishing new AI safety and transparency obligations for certain AI developers.

In September 2024, Governor Newsom **vetoed** an AI safety bill while still recognizing the need to protect the public from AI-related threats. Just over a year later, **TFAIA** aims to regulate AI developers and enhance AI safety, while simultaneously allowing California’s technology sector to continue to innovate.

Background

On September 29, Governor Newsom signed TFAIA, creating a regulatory framework for “frontier developers,” defined as persons or entities that have trained, or begun training, any foundation model trained with more than 10^{26} integer or floating-point operations. Notably, the calculation includes all computing power used for initial training, fine-tuning, reinforcement learning, or other material modifications.

Transparency Obligations for Frontier Developers

TFAIA subjects “large frontier developers” — any frontier developer that, together with its affiliates,

has combined annual gross revenues exceeding \$500 million — to robust transparency requirements.

- **Website Publication of AI Framework:** Large frontier developers must publicly publish an AI framework applicable to their frontier models. The framework must set out frontier developers' approach to incorporating international and national standards and industry best practices, identifying and mitigating catastrophic risks, use of third-party assessments, processes to update the framework as needed, security of unreleased model weights, and internal governance to manage and review risks before deploying or using frontier models.
- **Annual Review:** AI frameworks must be reviewed at least annually, and any updates must be clearly and conspicuously published along with justification of any material modifications within 30 days.
- **Transparency Report:** Before or at the time of deploying new frontier models or materially modified versions of prior models, large frontier developers must clearly and conspicuously publish a transparency report detailing intended uses, generally applicable restrictions or conditions of use, and a risk assessment.

Safety-Related Reporting Obligations

TFAIA directs the Office of Emergency Services to establish a mechanism for frontier developers or members of the general public to report a “critical safety incident.”^[1] Frontier developers must report such incidents to the Office of Emergency Services within 15 days of discovery when they pertain to one or more of the developer’s frontier models. Beginning January 1, 2027, the Office will publish an annual report that includes anonymized, aggregated information about critical safety incident reports from the prior year.

Whistleblower Protections

TFAIA protects covered employees who disclose significant health and safety risks posed by frontier models, including “catastrophic risk” as specified under the Act, arising from developing, storing, using, or deploying a foundation model. Large frontier developers must implement an internal process that enables employees to report, anonymously and in good faith, significant risks to public health or safety or legal violations, and must provide monthly updates to whistleblowers on the status and outcome of the investigation.

Encouraging Innovation

TFAIA creates a consortium within the Government Operations Agency that is charged with creating a public computer cluster, “CalCompute,” to support the development of safe, ethical, equitable, and sustainable AI by fostering research and innovation.

Annual Updates

The California Department of Technology must annually recommend updates to the law, where appropriate, taking into account input from multiple stakeholders, technological advancements, and international standards.

Key Takeaways

AI developers should promptly assess whether they qualify as a “frontier developer” or “large frontier developer” to determine applicable obligations and should implement a plan for compliance with the law’s requirements likely to go into effect January 1, 2026.

- **AI Framework and Transparency Reports:** Large frontier developers should draft a public-facing framework covering standards alignment, catastrophic risk identification or mitigation, third-party assessments, model-weight security, and pre-deployment governance. They should also ensure they maintain all information needed to publish transparency reports. Additionally, large frontier developers should schedule at least annual framework reviews and a process to publish material changes within 30 days.
- **Incident Reporting:** Implement processes to detect and document critical safety incidents and establish a mechanism for reporting the incidents to the Office of Emergency Services.
- **Whistleblower Notice and Reporting:** Post notice of employees’ rights under TFAIA within the workplace and directly notify employees of their rights at least once per year. Create a mechanism for whistleblowers to internally and anonymously report catastrophic risks and receive monthly status updates.

TFAIA creates significant penalties for non-compliant AI developers. Frontier developers that fail to publish or transmit a document required under TFAIA, or that do not report a critical safety incident to the Office of Emergency Services face a civil penalty of up to \$1 million per violation. Additionally, TFAIA creates a civil action, enforceable by the Attorney General’s office, against AI developers for non-compliance related to whistleblower protections.

[1] A “critical safety incident” is defined as including any of the following:

1. Unauthorized access to, modification of, or exfiltration of the model weights of a frontier model that results in death or bodily injury.
2. Harm resulting from the materialization of a catastrophic risk.
3. Loss of control of a frontier model causing death or bodily injury.
4. A frontier model that uses deceptive techniques against the frontier developer to subvert the controls or monitoring of its frontier developer outside of the context of an evaluation designed to elicit this behavior and in a manner that demonstrates materially increased catastrophic risk.

Contacts



Matthew Berlin

PARTNER AND AI, METAVERSE & BLOCKCHAIN
INDUSTRY GROUP CO-LEADER



Natasha Weis

ASSOCIATE

Related Industries

AI, Metaverse & Blockchain

Continue Reading

AI LAW BLOG

Judge Pauses Anthropic's \$1.5 Billion Copyright Settlement

SEPTEMBER 17, 2025 | MATTHEW L. FINKELSTEIN*, DANIELLE W. BULGER, NATASHA WEIS

AI LAW BLOG

Patenting AI: Proliferation of AI Technologies Across Diverse Industry Sectors

JULY 9, 2025 | MICHAEL FAINBERG, MOHAMMAD ZARYAB

All Insights from AI Law Blog

NY Class Action Filed Against Nike Over Alleged NFT ‘Rug Pull’ and Unregistered Securities

MAY 20, 2025 | ANTHONY V. LUPO, EMILY B. LEWIS, ALEXIS MOZELESKI