CALIFORNIA'S NEW AI SAFETY LAW

WHAT EVERY DEVELOPER NEEDS TO KNOW

by **John Eden** | Counsel

OVERVIEW OF THE NEW LAW

California has taken the lead in setting a regulatory floor for artificial intelligence. On September 29, 2025, Governor Gavin Newsom signed into law **Senate Bill 53**, the *Transparency in Frontier Artificial Intelligence Act* ("**TFAIA**"). This is the first comprehensive state law in the U.S. to require transparency, safety, and accountability from developers of advanced AI models and platforms.

The law was enacted after Congress failed to adopt a national standard and amid growing concern about catastrophic AI risks. In contrast to the federal government's light-touch approach (which involved a 10-year moratorium on the adoption of broad-based AI laws at the state level), California is sending a clear signal: *Companies that build or deploy powerful AI systems will be expected to manage risks, disclose safety practices, and protect employees who raise legitimate concerns*.

For entrepreneurs and legal professionals, the message is straightforward. Even if your company is not yet within the statute's reach, TFAIA sets a new benchmark that could soon become the de facto national standard for AI governance.

THE IMPACT OF TFAIA ON LARGE FRONTIER DEVELOPERS

The law focuses on developers of so-called "frontier" AI models, defined by their computational scale, potential for broad use, and material impact on the AI ecosystem itself.

What exactly is a frontier model? A frontier model is one trained using more than 10^{26} integer or floating-point operations—a threshold intended to capture powerful foundation models that can be used to power a wide range of consumer and enterprise products and services. A large frontier developer is any company (and its affiliates) with annual gross revenues exceeding $500 million in the previous calendar year.

These developers face the most extensive obligations under the law, including publishing a "***Frontier AI Framework***". This is a public document outlining how the company identifies and mitigates "***Catastrophic Risks***," defined as follows:

> *A material risk that a developer's creation, storage, use or deployment of a foundational model could materially contribute to the death of or serious injury to more than fifty (50) people, or at least $1 billion USD in property damage wherein the foundational model does one or more of the following: (1) provides assistance in the creation or release of a chemical, biological, radiological or nuclear weapon, (2) engages in activity without meaningful human oversight that constitutes or cyberattack or, if committed by a human, would be classified as murder, assault, extortion or theft; and/or (3) evades the control of its frontier developer or user.*
>
> A compliant Frontier AI Framework must explain:

a. How the company aligns with recognized national or international safety standards;
b. How it determines whether a model could create "catastrophic risk," such as aiding in the development of weapons or enabling large-scale cyberattacks;
c. How third-party experts or auditors are used to test safety measures; and
d. What cybersecurity and internal governance mechanisms protect model weights and ensure accountability.

Failure to comply can result in civil penalties up to $1 million per violation, enforced by the California Attorney General.

CORE PROVISIONS: SAFETY REPORTING AND WHISTLEBLOWER PROTECTIONS

The TFAIA creates two new compliance pillars: (i) safety incident reporting, and (ii) whistleblower protections.

Safety Reporting. Developers must report critical safety incidents to the California Office of Emergency Services ("**COES**"). These include:

a. Unauthorized access to model weights that leads to harm or injury;
b. Harm that is causes by the emergence of an event or series of events that qualifies as a catastrophic risk;
c. Loss of control of an AI model that causes death or serious bodily harm; and/or
d. A model autonomously engaging in deceptive or criminal behavior that heightens catastrophic risk.

Reports must be submitted to COES within 15 days of discovery, or within 24 hours if there is an imminent threat to life or safety. Although the reports are confidential to protect trade secrets, COES will publish anonymized annual

summaries to ensure that policymakers and the general public are well-informed.

Whistleblower Protections. The new law also gives employees responsible for AI risk management strong legal protections when reporting safety violations. Large frontier developers must provide anonymous internal reporting channels for safety concerns, update whistleblowers monthly on the status of their disclosures, and avoid any form of retaliation. In cases where retaliation is alleged, the burden of proof automatically shifts to the employer. Moreover, employees can seek injunctive relief and attorney's fees if the TFAIA's whistleblower protections are violated.

DOES THE TFAIA APPLY TO YOUR COMPANY?

Not every AI company falls within the law's scope. But the thresholds are low enough that many well-funded or rapidly scaling developers could indeed be covered in the near future.

Under the current thresholds, you need to ask these questions:

1. ***Compute Threshold***: Does your model training exceed 10^{26} floating-point operations?
2. ***Revenue Threshold***: Did your company and its affiliates generate at least $500 million USD in gross revenue last year?
3. ***Model Type***: Are you developing or deploying foundational or general-purpose AI models that could enable high-risk applications?

If you answer yes to any of these—or plan to in upcoming training cycles—you should begin preparing a compliance roadmap immediately. Even smaller developers should expect contractual flow-down obligations from partners, vendors, or clients who are covered entities under TFAIA.

HOW SHOULD YOU PREPARE FOR COMPLIANCE?

The TFAIA takes effect January 1, 2026, but preparation should begin now. Covered entities and companies that want to adopt best practices proactively should focus on two immediate actions:

1. ***Develop Incident Response Policies***: Establish a formal AI safety and incident response protocol that enables quick identification, assessment, and reporting of critical safety incidents. This should include defined escalation paths to senior leadership and legal, clear criteria for what qualifies as a "critical incident," coordination between technical, legal, and communications teams, and a process for timely reporting to COES if required.
2. ***Strengthen Internal Whistleblower and Governance Policies:*** Review and update HR policies, employment contracts, and NDAs to reflect new whistleblower protections. Employees should know (i) how to report concerns safely and anonymously, (ii) what protections they have under California law, and (iii) that retaliation is prohibited and carries serious consequences for an offending company.

FINAL THOUGHTS

California's TFAIA is much more than a state law: ***It's a preview of the next phase of AI regulation in the United States***.

For companies building advanced AI systems, compliance will require more than technical documentation; it will demand a new governance mindset that values safety, transparency, and ensuring that employee feedback is truly valued and taken seriously.

BurgherGray is a dynamic corporate boutique law firm comprised of highly experienced and diverse attorneys, most of whom have honed their skills practicing at large corporate law firms and in-house legal departments of large corporations and governmental agencies. The firm counsels clients ranging from emerging companies to Fortune 100 enterprises on a range of complex business litigation, government and internal investigations and transactional matters, including finance, M&A, securities regulation and corporate and commercial transactions. The firm is a member of NAMWOLF, a member of NMSDC, and is certified as an MBE by the City and State of New York and the City of Chicago and State of Illinois.

*To learn more about California's TFAIA, please contact John Eden by email (**jeden@burghergray.com**) or phone (**415.567.8025**).*

BURGHERGRAY LLP
ATTORNEYS AT LAW

www.burghergray.com

1350 Broadway | Suite 1510
New York, NY 10018
T: 646.513.3231 | F: 646.561.9866
info@burghergray.com