# Compliance in a Global AI Market: Examining the Overlaps Between California's SB 53 and the EU AI Act

Authored by: Beth George, Brock Dahl, Christine Chong, Vera Ibes, Joseph Mason, and Madeline Cimino

On September 29, Governor Newson signed into law Senate Bill 53, the "Transparency in Frontier Artificial Intelligence Act," ("TFAIA"), which marks a significant development in U.S. regulatory requirements pertaining to AI. Paring back from last year's more ambitious bill, SB 1047, this new legislation is intended to significantly expand transparency for key categories of large AI models and companies, termed "frontier models" and "large frontier developers," and to address the potential for catastrophic risks from such systems. The law establishes a framework for publicly disclosing significant model updates, publishing safety protocols, reporting critical safety incidents to public authorities, and more.

Companies already building their AI governance models around the EU AI Act should take special note. California's new law has a much narrower applicability than the EU AI Act - it applies to models over a certain size (a foundation model that was trained using a quantity of computing power greater than 10^26 integer or floating-point operations) rather than categorizing models by risk definitions. Nonetheless, both regimes feature similar requirements for risk management, documentation, and incident reporting for covered entities. Moreover, while the signing of TRAIA is introducing new requirements on AI companies domestically, a contrasting discussion is underway in Brussels regarding a potential "pause" or "stop the clock" on certain elements of the EU AI Act, to reduce the administrative burden for companies. Thus, companies must assess the degree to which their models are subject to both regimes, and where necessary, develop strategies for ongoing compliance with overlapping requirements.

### Covered Entities and Models

The TFAIA is designed to apply to a select group of models and their developers, defined by key criteria. The TFAIA's scope is defined by two primary terms: "**frontier model**" and "**large frontier developer**."

- The law defines a **"frontier model"** as a foundational artificial intelligence model that was trained using a quantity of computing power greater than a computational capacity exceeding 10^26 integer or floating-point operations.
- The law is also aimed at regulating **"large frontier developers"** - defined as a frontier developer whose collective annual gross revenues, together with its affiliates, exceeded $500 million in the preceding calendar year.

The law's use of a dual threshold for both technical capability and revenue is intended to apply the requirements to a limited number of developers and match regulatory oversight to the scale of potential risk. Comparatively, the EU AI

Act's regulations extend to a much broader range of companies and AI systems than the TFAIA. With the EU AI Act, the most comprehensive obligations focus on providers of high-risk AI systems and General Purpose AI ("GPAI") models, particularly those with systemic risk. The EU AI Act obligations for GPAI models apply to those trained using 10^25 integer or floating-point operations, capturing a much wider number of GPAI models than the TFAIA.

## Core Obligations: Transparency and Safety Frameworks

The Frontier AI Framework

A core requirement is that large frontier developers must write, implement, and clearly and conspicuously publish on their website a frontier AI framework. Each frontier model must have an applicable framework, describing the developer's approach to key governance areas, including:

- Standards and Practices: Incorporating national standards, international standards, and industry-consensus best practices.
- Risk Assessment and Mitigation: Defining thresholds used to identify and assess capabilities that could pose a catastrophic risk and detailing the relevant mitigations.
- Internal Governance: Instituting internal practices to ensure compliance, reviewing assessments, mitigations, and blocking unauthorized modifications.
- Third-Party Involvement: Detailing the use of third parties to assess potential catastrophic risks and the effectiveness of mitigation strategies.

This published framework is intended to serve as a definitive reference point for the developer's internal protocols. Notably, a failure to comply with its own frontier AI framework can subject a large frontier developer to civil penalties enforceable by the Attorney General. Developers are also required to review and update the framework at least once per year and must publish the modified framework along with a justification within 30 days of making any material modification.

Transparency Reports and Disclosures

In addition to the overarching framework, transparency reports are required at the time of deployment of new covered models. Before, or concurrently with, deploying a new or substantially modified frontier model, the developer must publish a transparency report. For large frontier developers, this report must include summaries of the catastrophic risk assessments conducted and the results of those assessments, detailing the extent of third-party evaluator involvement.

The law allows developers to make redactions to these published documents (both the framework and the reports) if necessary to protect trade secrets, cybersecurity, public safety, or national security. However, any redaction must be accompanied by a justification, and the redacted information must be retained for five years.

Risk Assessment and Incident Reporting

The law places a strong emphasis on mandatory reporting of actual and potential high-severity incidents. The law defines "catastrophic risk" as a foreseeable and material risk that a model's deployment will materially contribute to the death or serious injury of more than 50 people or cause more than $1 billion in damage to property from a single incident. This definition explicitly covers risks such as providing expert assistance in the creation of chemical, biological, or nuclear weapons, large-scale cyberattacks, or a model evading human control.

Developers are required to report any "critical safety incident" to the California Office of Emergency Services (OES) within 15 days of discovery. The definition of a critical safety incident includes unauthorized access to model weights that results in injury, loss of control causing injury, or a model using deceptive techniques to subvert internal controls in a manner that demonstrates materially increased catastrophic risk. Crucially, if a frontier developer discovers that a critical safety incident poses an imminent risk of death or serious physical injury, disclosure must be made to the appropriate law enforcement or public safety authority within 24 hours.

Comparison with EU AI Act Obligations

While both the EU AI Act and the TFAIA require formal risk management and documentation, their exact requirements differ. While imposing certain publication obligations, the EU AI Act primarily focuses on a highly prescriptive, compliance-heavy framework with required regulatory submissions. Comparatively, California's new law requires large frontier developers to publicly publish the higher-level Frontier AI Framework and transparency report, that primarily focus on communicating the company's approach to mitigating catastrophic risks and informing users about the intended use of the frontier model.

For incident reporting requirements, the EU AI Act and the TFAIA differ widely in the scope of the incidents covered but impose similar timing requirements for the broadest category of incidents. Under the TFAIA, developers must report "critical safety incidents" to the California Office of Emergency Services with time-bound deadlines of 15 days, or 24 hours for imminent risks. The EU AI Act requires providers of high-risk systems to report "serious incidents," which has a much broader definition that includes harm to fundamental rights, to national market surveillance authorities "immediately after the provider has established a causal link between the AI system and the serious incident or the reasonable likelihood of such a link," but no later than 15 days after the company becomes aware of the serious incident.

## Penalties and Whistleblower Protections

Financial Penalties for Noncompliance

The TFAIA imposes civil penalties, to be brought only by the California Attorney General, for large frontier developers who:

- Fail to publish or transmit a compliant document;
- Make a materially false or misleading statement about catastrophic risk or framework compliance;
- Fail to report an incident as required; or
- Fail to comply with its own frontier AI framework.

The maximum fine for each violation is $1,000,000.

Whistleblower Protections

The TFAIA introduces significant whistleblower protections for "covered employees," defined as those responsible for assessing, managing, or addressing the risk of critical safety incidents. A frontier developer is prohibited from retaliating against a covered employee who discloses information to the Attorney General, federal authorities, or others, provided they have reasonable cause to believe the information indicates a specific and substantial danger related to catastrophic risk or a violation of the TFAIA. Large frontier developers must also implement an internal process for covered employees to anonymously disclose concerns.

## Key Takeaways

The signing of the TFAIA establishes a new legal framework that marks a notable shift from a system of voluntary industry best practices to one of mandatory, state-mandated compliance. Companies already compliant with the EU AI Act's rigorous requirement relating to GPAI models will be well-positioned to address the transparency and reporting obligations mandated by the TFAIA, but should take special note of the distinctions, as well. As the EU AI Act requirements in relation to high-risk systems come into force over the next two years, companies should ensure their compliance approach incorporates both the TFAIA's requirements, and the EU's more prescriptive requirements for technical documentation and risk classification.