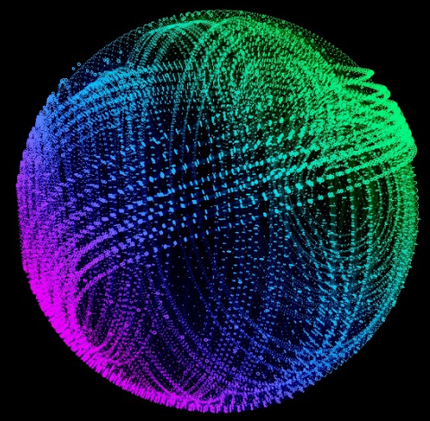
Deloitte.

5x5 series: Insights and actions

Insights and actions on *SB53: Transparency in Frontier Artificial Intelligence Act*



SB53, signed into California law on September 29, 2025, is the first state law in the US to regulate developers of frontier artificial intelligence (AI) models[1,2]. Taking effect on January 1, 2026, it includes safety guardrails, incident reporting, whistleblower protections, and other compliance requirements, with penalties for non-compliance.

5 insights you should know

1 **Transparency disclosure of a frontier AI framework**
Large frontier developers must publish disclosure of a frontier AI framework on their website, stating how they have incorporated leading standards for AI governance and risk management into their frontier models.

2 **Disclosure of individual model risk and mitigations**
Before—or concurrently with—deployment of a new or significantly modified frontier AI model, a frontier developer is required to include in their transparency report summaries of the catastrophic risk assessments they have conducted, the results of these assessments, information about the involvement of third-party evaluators, and a description of any actions taken to meet the requirements of the frontier AI framework for that particular model.

3 **Safety and incident reporting**
The law additionally requires frontier developers to report "critical safety incidents" to California's Office of Emergency Services within 15 days of discovery. Critical safety incidents include (1) unauthorized access to, modification of exfiltration of, the model weights of a frontier model that result in death or bodily injury; (2) harm resulting from the materialization of catastrophic risk; or (3) loss of control of a frontier model causing death or bodily injury.

4 **Whistleblower channels and protections**
The law requires a large frontier developer to provide an internal process through which a covered employee may anonymously disclose information if the covered employee believes in good faith that the large frontier developer's activities present specific and substantial danger to public health or safety. The law additionally requires the California Office of Emergency Services to create a mechanism to be used by a frontier developer or a member of the public to report a critical safety incident.

5 **SB53 joins other international AI regulation**
SB53 joins international AI regulations like the EU *Artificial Intelligence Act*[3], as well as other state regulations that govern specific risks within AI systems, including the development of deepfakes, protection of intellectual property, and child safety.

5 actions you can take

1 **Refresh and assess internal frameworks**
Compare AI policies to industry-leading frameworks, such as those provided by the *National Institute of Standards and Technology (NIST) AI 600-1* or *International International Organization for Standardization (ISO) 42001:2023,* and consider an internal assessment to determine whether the controls and governance processes designed to support these frameworks are operating effectively.

2 **Publish transparency reports on frontier AI framework, governance, and mitigations**
Publish transparency reports detailing how your organization is applying standards into your frontier AI framework; defining and assessing thresholds on capabilities that could pose a catastrophic risk, applying mitigations to address these risks; the cybersecurity practices to secure unreleased model weights; identifying and responding to critical safety incidents; internal governance practices; and its assessment and management of catastrophic risk resulting from internal use of the frontier models.

3 **Understand the types of "critical safety incidents" that may apply to your frontier model**
Work with frontier model developers and risk teams to understand the types of critical safety incidents that can occur based on the model's capabilities. Consider red teaming and other testing to assess model guardrails and develop a risk reporting framework that clarifies critical safety incidents in the context of SB 53's definition.

4 **Update whistleblower programs and communications**
Update communications policies to clarify that whistleblower hotlines and other intake channels can be utilized to disclose information about artificial intelligence systems and related processes (including the teams developing these tools and supporting technologies). Provide resources to the teams that are supporting these channels to interpret how to inquire and investigate these concerns.

5 **Build an AI risk management program that can scale across multiple regulations and requirements**
There could be overlapping requirements across AI regulations – including obligations to inventory AI systems, provide transparency, and consumer protection. Organizations should develop a scalable program that inventories, analyzes, and assesses compliance against common requirements.

1: Bill Text - SB-53 Artificial intelligence models: large developers.
2: Section 22757.10: "Frontier models" are defined as those that have been trained using a quantity of computing power greater than 10^26 integer or floating-point operations. "Frontier developers" refer to those who have trained, or initiated the training of, a frontier model; and "large frontier developers" are those with annual gross revenues in excess of $500,000,000 in the preceding calendar year.
3: The AI Act Explorer | EU Artificial Intelligence Act

Connect with us

Rich Tumber
Principal
Deloitte & Touche LLP
ritumber@deloitte.com

Cliff Goss
Partner
Deloitte & Touche LLP
cgoss@deloitte.com

Brendan Maggiore
Senior Manager
Deloitte Transactions and Business Analytics LLP
bmaggiore@deloitte.com