Davis Polk

New York joins California in regulating development of frontier AI models

May 20, 2026 | Client Update | 9-minute read

Beginning January 1, 2027, the New York RAISE (Responsible AI Safety and Education) Act will require developers of frontier AI models to file disclosures, publish detailed risk management documents and report safety incidents.

Introduction

On March 27, 2026, New York Governor Kathy Hochul signed the final version of the RAISE Act, aiming to establish "a strong and sensible standard for frontier AI safety, holding the biggest developers accountable for their safety and transparency protocols." The final version will go into effect on January 1, 2027, creating two-tiered requirements for developers of "frontier" AI models. The Act imposes the most significant obligations on "large frontier developers," including to implement, publish, and adhere to frameworks governing the development and assessment of their frontier models, and imposes additional requirements on all frontier developers, such as obligations to publish transparency reports about their models and to report critical safety incidents to the New York Department of Financial Services (DFS). The Act also calls for an (as-yet-unnamed) office within the DFS to administer the Act and enforce certain provisions against large frontier developers, as well as to approve federal laws or guidance as alternative standards for incident reporting. The New York Attorney General will also be empowered to enforce certain obligations against large frontier developers.

A condition of Governor Hochul's approval of the initial legislation on December 19, 2025 was the New York Assembly's formal passage of agreed-upon chapter amendments to the Act that bring the RAISE Act more in line with California's Transparency in Frontier Artificial Intelligence Act (TFAIA) (summarized here). For example, both laws apply to large developers of frontier AI models and require, among other things, publication of AI frameworks explaining the developer's approach to establishing "catastrophic risk" thresholds and critical safety incidents, what mitigations and cybersecurity practices they employ, and details about internal governance, including incident response. Notable differences include a 72-hour incident reporting timeline under the RAISE Act compared with a 15-day timeline under TFAIA, a grant of rulemaking authority to a state regulatory body, and the absence of whistleblower protections. The original act, signed on December 19, 2025, also came approximately one week after President Trump issued Executive Order 14365 setting forth a series of directives, including potential lawsuits against states, aimed at checking "the most onerous and excessive" state AI laws. Although the final version of the RAISE Act tempers its requirements to more closely track TFAIA, it remains to be seen how federal policy will shape state AI laws in New York and elsewhere.

Applicability

The RAISE Act applies to "frontier developers" who develop, deploy, or operate in New York one or more "frontier models," defined as models trained using a quantity of computing power greater than 10°26 integer or floating-point operations. This threshold is identical to that in TFAIA, and like TFAIA, this threshold includes compute used in original training runs as well as subsequent training or fine-tuning applied to a preceding foundation model.

While certain RAISE Act requirements apply to all frontier developers, the Act reserves the most onerous requirements—and related civil enforcement—for "large frontier developers," who have gross annual revenues exceeding

$500 million in the last calendar year (as does California's TFAIA).

Transparency requirements

Frontier AI frameworks

As in TFAIA, the RAISE Act requires large frontier developers to "write, implement, comply with, and clearly and conspicuously publish" on their website a "frontier AI framework" describing how the developer approaches specific aspects of developing frontier models. Large frontier developers must review and update their frameworks annually and republish the modified version (with a justification for the modification) in the event of "material" changes. Large frontier developers are also prohibited from making materially false or misleading statements about their implementation of, or compliance with, their frameworks, unless made in good faith and reasonable under the circumstances.

As for TFAIA, published frontier AI frameworks must describe in detail how the large frontier developer:

- Incorporates national and international standards, as well as industry best practices;
- Defines thresholds used to assess whether its frontier model could present a catastrophic risk;
- Applies mitigations to address these potential risks, and how it assesses the adequacy of mitigations as part of any decision to deploy or use such a model, even internally;
- Uses third parties to assess the potential for catastrophic risks and the effectiveness of related mitigations;
- Revisits and updates its frontier AI framework;
- Secures its unreleased model weights from unauthorized modification or transfer;
- Identifies and responds to critical safety incidents; and
- Institutes governance practices to ensure compliance.

The Act defines "catastrophic risk" similarly to TFAIA: a foreseeable and material risk of death or serious injury of more than 50 people, or at least 1 billion dollars of damages, caused or materially enabled by a frontier developer's use, storage, or release of a frontier model, through either:

- Expert-level assistance in the creation or use of chemical, biological, radiological or nuclear weapons;
- Conduct by an AI model, without meaningful human intervention, that would constitute a cyberattack, murder, assault, extortion, or theft (including theft by false pretense) if committed by a human; or
- Evading the control of the developer or user.

"Catastrophic risk" does not include risks related to the frontier model outputting information that is otherwise publicly accessible; lawful activity of the federal government; or harm caused in combination with software, so long as the frontier model did not materially contribute to the harm.

Public transparency reports

The RAISE Act also creates transparency obligations for all developers of frontier AI models, with heightened requirements for large frontier developers. When any frontier developer deploys a new or substantially modified frontier model, it must clearly and conspicuously publish on its website a transparency report containing, among other things, a mechanism that enables the public to contact the frontier developer, the intended uses of the frontier model, and any generally applicable restrictions or conditions on uses of the frontier model. Similar to the prohibition on certain false or misleading statements related to frontier AI frameworks (which only applies to large frontier developers), all frontier developers are prohibited from making materially false or misleading statements regarding catastrophic risk from their frontier models or their management of that risk, unless made in good faith and reasonable under the circumstances.

Large frontier developers must additionally include summaries of assessments regarding catastrophic risk conducted pursuant to their frontier AI framework and the results, the role of third parties in those assessments, and other steps taken to comply with the developer's frontier AI framework. Such developers must also submit summaries of any assessments of catastrophic risk resulting from internal use of their frontier models every three months to the DFS, unless the DFS agrees to another reasonable schedule.

The Act permits redactions that are "necessary" to protect trade secrets, cybersecurity, public safety, national security, or to comply with federal or state law, but published versions must describe the character and justification of any redactions and developers must retain unredacted information for five years.

Disclosure statements

Prior to developing, deploying, or operating a frontier model, large frontier developers must file a disclosure statement with the DFS and pay a pro rata share of expenses related to administration of the Act. These disclosure statements must provide the developer's contact information, including for inquiries from the DFS or other governmental entities, and identify the large frontier developer's current and recent ownership. The DFS can require additional information, and disclosure statements must be renewed every two years or whenever there is a material change in the information provided, such as a change in ownership. Large frontier developers who fail to file a disclosure statement or pay their share may be subject to daily civil penalties after notice and a hearing. TFAIA does not require these disclosure statements.

Critical safety incident reporting

The reporting obligations require large developers to report any critical safety incidents to the DFS within 72 hours, a shorter window than TFAIA's 15-day requirement—and one of the Act's most notable deviations, though the reporting criteria are otherwise the same. This timeline is reduced to 24 hours where a frontier developer discovers that a critical safety incident poses an imminent risk of death or serious physical injury. "Critical safety incidents" include:

- Unauthorized access to, modification of, or exfiltration of a frontier model's model weights that results in death or bodily injury;
- A catastrophic risk materializing into harm;
- Loss of control of a frontier model causing death or bodily injury; or
- A frontier model using deceptive techniques against the developer, outside of deception-focused testing, to subvert controls or monitoring in a manner that materially increases catastrophic risks.

The Act also allows for members of the public to file critical safety incident reports, although the DFS has discretion to review them. The DFS may transmit reports of critical safety incidents, or summaries of any assessments of catastrophic risk to other governmental entities, including the New York Attorney General.

Federal law alternatives for incident reporting

The DFS may adopt regulations that allow frontier developers to use compliance with federal laws, regulations, or guidance documents to meet their RAISE Act incident reporting obligations. The DFS can only promulgate and maintain such regulations for federal incident-reporting standards that are substantially equivalent to, or stricter than, the RAISE Act's requirement. Frontier developers must notify the DFS if they intend to use federal reporting requirements in lieu of making the RAISE Act's required reports, and they must send copies of any critical safety incident reports required by the federal standards. Notably, though New York appears to have sought to help future-proof the RAISE Act by allowing for compliance with federal law, no such federal incident-reporting standards exist to date.

Oversight and enforcement

The Act imposes liability on large frontier developers for failure to meet the Act's transparency requirements, to report an incident as required, or to comply with frontier AI frameworks. Large frontier developers can also be held liable for false or misleading statements about their implementation of, or compliance with, their frontier AI frameworks. Notably, the Act does not provide for enforcement against frontier developers related to the prohibition on materially false or misleading statements regarding catastrophic risk from their frontier models and their management thereof.

The RAISE Act empowers the New York Attorney General to seek civil penalties of up to $1 million for first violations, and up to $3 million for subsequent violations—TFAIA also imposes fines of up to $1 million per violation, but does not distinguish between first and subsequent violations. The Act expressly disclaims a private right of action.

As noted above, the RAISE Act also creates a new office within the DFS that will receive disclosure statements, review risk assessments and incident reports, and make annual reports to the Governor and Legislature. In addition to generally

overseeing implementation of the RAISE Act, the DFS will be empowered with rulemaking authority to adopt regulations subject to notice and comment requirements. The DFS will also assess the pro rata fees on large frontier developers and enforce against large frontier developers that fail to file required disclosures or pay their pro rata shares. Failure to keep a current and accurate disclosure filed with the DFS, or failure to timely pay any assessment, can result in civil penalties equal to $1,000 per day of noncompliance, plus the amount of assessments owed, as well as fees and costs.

Key takeaways

- Legal and compliance functions should socialize RAISE Act requirements with business and product development counterparts early, given the Act's requirement that documents be published alongside new frontier model releases.
- Covered developers should seek to leverage compliance protocols developed for TFAIA for the RAISE Act as well, given their many similarities.
- Companies that develop frontier AI models should start developing frontier AI frameworks, utilizing industry standards and third-party assessment where appropriate to bolster common defensibility across the increasing patchwork of AI laws and regulations.
- Companies should assess their safety incident reporting protocols against the RAISE Act, which requires more immediate reporting of safety incidents than California's TFAIA.
- Companies should closely monitor for rulemaking activity by the DFS to implement the RAISE Act.

If you have any questions regarding the matters covered in this publication, please reach out to any of the lawyers listed below or your usual Davis Polk contact.

Matthew J. Bacal
+1 212 450 4790
matthew.bacal@davispolk.com

David I. Feinstein
+1 212 450 3293
david.feinstein@davispolk.com

James W. Haldin
+1 212 450 4059
james.haldin@davispolk.com

David Lisson
+1 650 752 2013
david.lisson@davispolk.com

Howard Shelanski
+1 202 962 7060
howard.shelanski@davispolk.com

Maude Paquin
+1 212 450 3205
maude.paquin@davispolk.com