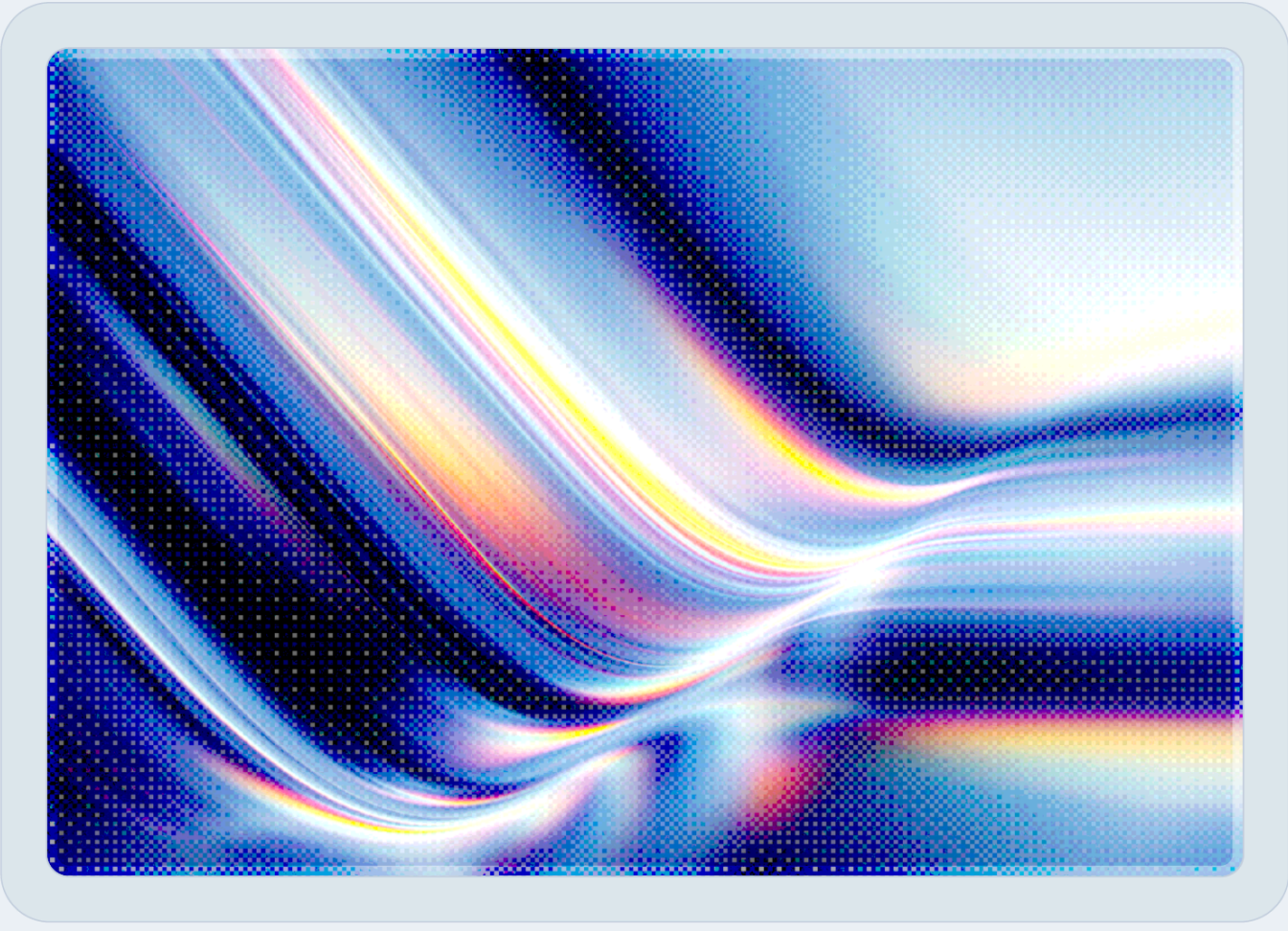
CSAI Foundation | Cloud Security Alliance

After the Midterms: Preparing for Frontier-Model Audits and 72-Hour Reporting

Security Implications and Guidance for Frontier Developers and the Enterprises That Depend on Them

2026-11-08

AI-assisted Rapid Research



CSAI | CSA cloud security alliance®

© 2026 Cloud Security Alliance. Some rights reserved.

You may download, store, display, view, print, redistribute, and link to this document in its original, unmodified form, provided that attribution to the Cloud Security Alliance is maintained and all trademark and copyright notices remain intact.

This document may not be modified or altered. You may quote portions of the document as permitted by the Fair Use provisions of the United States Copyright Act, provided that attribution is given to the Cloud Security Alliance.

This document may be shared on professional and social media platforms in its original form with attribution.

This document was generated with AI assistance and has not undergone official CSA review and approval processes.

Key Takeaways

The November 3, 2026 midterms gave Democrats control of both chambers of Congress, and the 120th Congress takes office in January 2027 [1][2]. Pending federal frontier-AI bills, including the bipartisan FRONTIER Act (H.R. 9925), will have to be reintroduced [1][4]. In CSA's assessment, the political conditions for mandatory independent audits and incident reporting at the federal level now look more favorable than they did before the election. Mandatory frontier oversight is advancing without Congress in any case. New York's amended RAISE Act requires covered developers to report critical safety incidents within 72 hours starting January 1, 2027 [5]. Illinois's Artificial Intelligence Safety Measures Act (SB 315) adds its own 72-hour reporting window and requires annual independent third-party audits beginning in 2028 [6].

The 72-hour clock starts when a developer determines that an incident occurred or learns facts that support a reasonable belief that one did. It does not wait for an investigation to conclude [5][6]. Several of the defined incident categories, such as loss of model control or a model deceiving its developer to evade monitoring, describe model behavior rather than system compromise. In CSA's assessment, most security operations centers are not yet built to detect them.

Mandatory audits also raise a security problem, not only a compliance one. Auditors may need privileged access to model weights, evaluation pipelines, and internal risk assessments, and California is studying whether verifiers should be embedded onsite inside frontier labs [9][10]. Most enterprises that deploy frontier models fall outside the direct scope of these laws, but they will likely feel the effects through vendor attestations and contractual incident notifications. Separately, a pending Senate bill would require AI providers holding Department of Defense agreements of $100 million or more to report national security incidents to DoD within 72 hours [11].

Background

For most of 2026, U.S. frontier-AI oversight developed along two separate paths. At the federal level, the administration favored voluntary commitments and challenged state regulation. A December 2025 executive order directed the Attorney General to establish an AI Litigation Task Force to challenge state AI laws considered inconsistent with national policy, and it called for a federal legislative framework that would preempt such laws [12]. The administration added a new element on September 29, 2026, when the White House released a voluntary accord with several leading AI developers. The accord recommends four layers

of control for frontier model developers. The third is an independent external auditor that assesses whether controls are followed and whether they work [8]. The accord is explicitly not legally binding. The published summary of it [8] describes no incident reporting commitment.

The states moved faster and imposed binding requirements. California's Transparency in Frontier Artificial Intelligence Act (SB 53) requires frontier developers to report critical safety incidents to the Office of Emergency Services within 15 days of discovery. When an incident poses an imminent risk of death or serious physical injury, the developer must notify an appropriate authority within 24 hours [13]. New York then tightened that model. A chapter amendment to the RAISE Act, signed March 27, 2026 and effective January 1, 2027, applies to frontier developers, defined by training compute above 10^26 operations. It also sets a large-developer tier for those with more than $500 million in annual revenue, including affiliates. Covered developers must report critical safety incidents within 72 hours to a new office within the New York Department of Financial Services (NYDFS), with a 24-hour window when there is imminent risk to life [5]. Penalties start at $1 million for an initial violation and rise to $3 million for subsequent violations [5].

Illinois followed on July 6, 2026, when Governor Pritzker signed SB 315. The law applies to large frontier developers with more than $500 million in prior-year revenue, including affiliates. It sets 72-hour and 24-hour reporting windows, with reports going to the Illinois Emergency Management Agency and Office of Homeland Security (IEMA-OHS) and the Attorney General. It also requires covered developers to retain independent third parties to audit compliance annually starting in 2028 [6]. The table below compares how the New York and Illinois laws define coverage. The revenue tests apply to the large-developer obligations, which are layered on top of the compute-based frontier developer definition.

Statute	Compute threshold	Revenue threshold	Notes
New York RAISE Act (amended)	More than 10^26 operations [5]	$500 million, inclusive of affiliates [5]	Excludes companies that only use or build on others' models [5]
Illinois SB 315	Applies to frontier models; see [6]	More than $500 million prior-year revenue, inclusive of affiliates [6]	Annual audits apply to large frontier developers from 2028 [6]

California has also begun building the infrastructure for third-party verification. SB 813, signed September 9, 2026, directs the state to develop standards for designating independent verification organizations [9] [10]. Executive Order N-9-26, signed September 18, directs state officials to consider three further steps. These are requiring large frontier developers to embed those organizations onsite in their laboratories, requiring an independently tested emergency shutdown mechanism, and updating the definition of critical

safety incidents "to include a range of loss-of-control incidents" [10]. Press coverage described this last change as covering systems that "appear to evade controls or operate beyond their intended boundaries" [9]. Recommendations are due November 16, 2026 [9][10].

Congress produced several competing proposals before the election. The FRONTIER Act, introduced July 23, 2026 by Representatives Trahan and Obernolte with a bipartisan group of cosponsors, would impose tiered requirements on developers above the 10^26 compute threshold, including risk management frameworks, independent audits, and incident reporting [3][4]. Secondary analysis of the bill indicates that the largest developers would retain independent verification organizations licensed under Department of Commerce rules. The same analysis indicates that states would be barred from imposing new substantive obligations in the covered transparency, audit, and reporting space [4]. In the Senate, Senators Banks and Gillibrand introduced the bipartisan Insider Threat Reporting and Security Guidance Act. It would require AI contractors with Department of Defense agreements worth $100 million or more to report national security incidents, such as theft of model weights, within 72 hours of discovery. Material vulnerabilities or concerning model behavior would have to be reported within seven days [11].

The election result changes the calculus for each of these efforts. Democrats won majorities in both chambers [1][2]. Reporting on the outcome notes three constraints. Bills not completed in the current Congress must be reintroduced in January, most contested Senate legislation still needs 60 votes to end debate, and the new majority lacks the two-thirds margin needed to override a veto [1]. Democrats have also proposed the GUARDRAILS Act to overturn the executive order targeting state AI regulation [1]. CSA's reading is that these conditions make aggressive federal preemption of state frontier laws less likely in 2027, while leaving room for a negotiated federal standard that includes audits and incident reporting. That is an inference about political dynamics, not a forecast of specific legislative outcomes.

Security Analysis

The statutes described above share a common design. They impose short reporting deadlines, define incidents in terms of model behavior and catastrophic harm, and increasingly rely on independent third parties to verify what developers claim. Each of these choices has security consequences that go beyond legal compliance. The subsections that follow cover four of them: when the reporting clock starts, how to manage divergent deadlines, the new attack surface that audits create, and the effects on organizations that deploy frontier models rather than build them.

The Reporting Clock Starts Before the Investigation Ends

In CSA's view, the most consequential operational detail in the New York and Illinois laws is what starts the 72-hour window. Both statutes start the clock when a developer determines that an incident occurred or learns facts establishing a reasonable belief that one did [5][6]. In practice, the window may open during triage, well before the root cause is known. Organizations that wait for investigative certainty before engaging legal and regulatory teams risk missing the deadline. Mature cyber incident programs are used to similar constraints. Existing regimes such as NYDFS's cybersecurity regulation already impose 72-hour notification, and the Cyber Incident Reporting for Critical Infrastructure Act (CIRCIA) will require 72-hour reporting of covered cyber incidents once CISA's final rule takes effect [15]. The difference is that the frontier statutes define incidents in terms of model behavior and catastrophic harm, not system compromise alone.

New York's definition of a critical safety incident shows the gap. It covers four categories [5]. The first is unauthorized access to or exfiltration of model weights that results in death or bodily injury. The second is harm resulting from a catastrophic risk materializing, and the third is loss of control of a frontier model that causes death or bodily injury. The fourth is a model using deceptive techniques against its developer, outside of designed evaluations, to subvert controls or monitoring in a way that shows materially increased catastrophic risk.

Only the first category resembles events that existing data-loss and insider-threat tooling is designed to catch. Even there, that tooling detects the precursor exfiltration, not the resulting harm that makes the event reportable. The second category depends on downstream harm signals that usually originate outside the developer's environment. The last two require something different. To detect that a model has subverted monitoring, a developer needs behavioral telemetry from training, evaluation, internal deployment, and agentic operation. It also needs analysts who can judge whether an anomaly reflects deliberate evasion or an ordinary failure. California's pending reconsideration of loss-of-control incidents would widen this category further [9][10]. In CSA's assessment, few security operations centers today have runbooks, detection content, or escalation paths for these events, because they were built around system compromise rather than model behavior.

Divergent Windows Require a Single Pipeline

Developers operating across jurisdictions face several overlapping clocks, recipients, and thresholds. The table below summarizes the main regimes relevant to U.S. frontier developers and their large AI contractors. Federal entries are pending bills or voluntary instruments, not enacted law. The EU entries are included for comparison. Frontier developers placing general-purpose AI models with systemic risk on the EU market are subject to Article 55, while Article 73 covers a different population, providers of high-risk AI systems.

Regime	Status (as of Nov. 8, 2026)	Standard window	Accelerated window	Recipient	Independent audit
California SB 53	In force	15 days from discovery [13]	24 hours, imminent risk to life [13]	Office of Emergency Services [13]	Under study (SB 813, EO N-9-26) [9][10]
New York RAISE Act (amended)	Effective Jan. 1, 2027 [5]	72 hours [5]	24 hours, imminent risk to life [5]	NYDFS frontier office [5]	Not mandated; third-party evaluator involvement disclosed [5]
Illinois SB 315	Phased; audits from 2028 [6]	72 hours [6]	24 hours, imminent risk [6]	IEMA-OHS and Attorney General [6]	Annual, starting 2028 [6]
FRONTIER Act (H.R. 9925)	Pending; must be reintroduced [1][4]	Not specified in public summaries [4]	Not specified	Not specified in public summaries; state AGs may opt in [4]	Licensed verifiers for largest developers [4]
Banks–Gillibrand DoD bill	Pending [11]	7 days, material vulnerabilities and concerning behavior [11]	72 hours, national security incidents [11]	Department of Defense [11]	Not described [11]
White House Accord	Voluntary [8]	None described [8]	None described	None described	Recommended external auditor [8]

Regime	Status (as of Nov. 8, 2026)	Standard window	Accelerated window	Recipient	Independent audit
EU AI Act, Art. 55 (GPAI with systemic risk)	Applicable to GPAI providers [20]	Without undue delay [20]; Code of Practice sets 2–15 day tiers by severity [22]	Shortest tiers for the most severe incidents [22]	AI Office [20]	External evaluation expected in some cases under the Code of Practice [22]
EU AI Act, Art. 73 (high-risk systems, for comparison)	Regulation in force; high-risk obligations postponed to Dec. 2, 2027 (Annex III) and Aug. 2, 2028 (Annex I) [21]	15 days [14]	10 days (death); 2 days (widespread or critical infrastructure) [14]	Market surveillance authorities [14]	Mostly self-assessed conformity

A statute-by-statute approach is likely to produce duplicate investigations, inconsistent narratives across regulators, and missed deadlines. A single intake and classification pipeline that applies the strictest applicable window first and then routes tailored notifications to each recipient is easier to defend and to audit. Inconsistent reports to different regulators about the same event may also create legal exposure in their own right, since New York penalizes materially false or misleading statements as well as failure to report [5].

Audits Expand the Attack Surface

Mandatory independent audits strengthen accountability, but they create new trust relationships around the most sensitive assets a frontier developer holds. An auditor assessing whether controls "work," as the White House accord puts it [8], may need access to model weights or the systems that handle them, along with evaluation harnesses, red-team results, internal catastrophic-risk assessments, and incident records. California's proposal to embed verification organizations onsite would place third-party personnel inside

lab environments on a continuing basis [9][10]. Each of these access paths is a potential exfiltration route. The auditor's own environment also becomes a target for adversaries seeking frontier capabilities or unreleased safety findings.

The audit market is also immature. An April 2026 analysis by AVERI observed that expertise is concentrated inside the companies being audited. It warned that a rush of demand could let unqualified vendors offer "rubber stamp" audits, and noted that access and data-protection standards for auditors are still being drafted [7]. The analysis recommended that audits begin by verifying compliance with each company's own published policies before moving to common standards [7], which is broadly the approach Illinois takes [6]. For security teams, this means the first audit cycles will likely test whether a developer's published frontier framework matches its actual practice. Controls that exist only on paper are likely to be among the most visible findings.

Transparency Creates Its Own Exposure

Published frontier frameworks, transparency reports, and redacted audit findings serve legitimate accountability goals, and some regulators expect public disclosure [5][7]. They also tell adversaries which risks a developer considers material, which mitigations are in place, and where evaluation coverage is thin. California's statute directs regulators to weigh trade secrets, cybersecurity, public safety, and national security when transmitting reports [13]. Even so, developers remain responsible for drafting disclosures that satisfy the law without publishing a roadmap for attack. This tension is likely to sharpen once Illinois requires summarized, appropriately redacted audit reports to be published [6], and potentially under federal law.

Downstream Effects for Deployers

Most enterprises that use frontier models through APIs or cloud platforms are not frontier developers under these statutes. New York's law, for example, does not cover companies that only use or build on others' models [5]. Even so, deployers are likely to see second-order effects. Vendor security questionnaires and contracts are likely to reference statutory incident categories more often. Deployers that observe anomalous model behavior in their own agentic deployments may also become the first source of the facts that start a provider's reporting clock. The Banks–Gillibrand bill does not reach deployers generally, since it targets AI contractors with large DoD agreements [11]. Organizations that both deploy frontier models and supply AI to DoD under such agreements should check whether they fall within its scope.

Recommendations

The recommendations below follow the timeline of the obligations themselves. New York's reporting duty begins in less than eight weeks, the first half of 2027 will bring federal bill reintroductions and California's verifier standards, and Illinois audits begin in 2028. Each phase builds on the one before it, so procedural groundwork laid now becomes the evidence base for later audits.

Immediate Actions (Before January 1, 2027)

Organizations that train frontier-scale models should confirm now whether they are covered by the California, New York, and Illinois statutes. That review should test the 10^26 compute threshold and, for large-developer obligations, the $500 million revenue threshold, including affiliate revenue, which both New York and Illinois count [5][6]. The New York effective date is less than eight weeks away, and the most important preparation is procedural rather than technical. Each covered developer should adopt a written clock-start protocol. It should define who may declare a reasonable belief that a critical safety incident occurred, how that determination is documented, and how legal, security, and safety teams are paged within hours rather than days.

Developers should also extend their incident taxonomy to cover model-behavior categories alongside traditional cyber events. A practical starting point is to add the four New York categories as named incident types, each with initial detection sources and an owner. Developers should then run at least one tabletop exercise that starts from an ambiguous signal, such as an internally deployed agent disabling its own logging, and walks through the 24-hour and 72-hour decision points. Enterprises deploying frontier models should review their contracts with model providers for incident notification terms. They should also define an internal channel for reporting anomalous model behavior to the provider, since that information may be material to the provider's statutory obligations.

Short-Term Mitigations (First Half of 2027)

Covered developers should build a single incident pipeline that classifies each event once, applies the strictest applicable deadline, and generates jurisdiction-specific notifications from a common record. The same record should serve as the evidence base for later audits, so that timelines do not have to be reconstructed after the fact.

Developers should treat audit readiness as an access-control design problem. They should define the minimum access each audit activity requires and prefer controlled evaluation environments to transferring weights or raw data. All auditor activity should be logged and reviewed, and auditor personnel should be

subject to the same insider-threat controls as employees with equivalent privileges. Contracts with auditors should set security requirements for the auditor's own environment, define breach notification duties, and govern how unredacted findings are handled.

Developers should also map their published frontier frameworks to an established control catalog, so that each claim has an owner, an operating control, and retained evidence. CSA's AI Controls Matrix and its auditing guidelines for model providers, discussed below, provide a structure for this mapping.

Strategic Considerations

The federal picture remains unsettled. As characterized in secondary analysis [4], the FRONTIER Act's preemption provision would bar new state obligations in the covered transparency, audit, and reporting space. Meanwhile, the incoming majority has proposed the GUARDRAILS Act to overturn the executive order targeting state AI regulation [1]. Organizations should design for the strictest applicable obligation, since a federal standard, if one passes, may arrive after the New York and Illinois obligations are already in force. Compliance teams should track three developments: California's November 16 recommendations on onsite verification, shutdown mechanisms, and incident definitions [9][10]; the reintroduction of federal bills in January; and the development of auditor licensing or designation standards in California and potentially at the Department of Commerce [4][10].

Over the longer term, several developments point the same way. Illinois has mandated audits, the White House accord recommends external auditors, California is writing verifier standards, and the EU expects external evaluation of systemic-risk models in some cases. Together they suggest that third-party assurance of frontier model controls is likely to become a more common expectation, although the mechanisms differ substantially. Developers that invest early in auditable controls, evidence retention, and secure auditor access are likely to be better positioned than those that treat each statute as a separate project.

CSA Resource Alignment

The most directly relevant CSA resource is the *AICM v1.1 Auditing Guidelines for Model Providers* [17]. It provides control-by-control verification procedures for AI model providers across the model lifecycle. That work closely matches what independent auditors are expected to do under Illinois SB 315 and under the verification models being considered in California and Congress. Early audit regimes are expected to test whether developers follow their own published frameworks [6][7]. Model providers can therefore use these guidelines to run an internal pre-audit, find controls that exist only in policy, and organize evidence in a form an external assessor can review.

The underlying *AI Controls Matrix (AICM) v1.1* [16] provides the control catalog against which frontier frameworks can be mapped. Developers can map their commitments on catastrophic-risk assessment, model-weight protection, evaluation integrity, and incident management to AICM control objectives. The result is a structure that can be maintained as state and federal definitions shift, which addresses the fragmentation described in "Divergent Windows Require a Single Pipeline."

For the 72-hour reporting challenge, CSA's *Cloud Incident Response Framework* [18] provides a foundation for incident detection, triage, escalation, and coordination in shared-responsibility environments. It is a reasonable starting point for extending incident taxonomies to model-behavior events. Organizations will still need to add AI-specific detection sources and decision criteria for categories such as loss of control or deceptive evasion.

Finally, *The Evolution of STAR: Introducing Continuous Auditing* [19] describes CSA's approach to continuous, evidence-driven assurance. Regulators are considering onsite verifiers and recurring audits for frontier developers. Continuous assurance practices from STAR offer a model for moving from point-in-time attestations to ongoing verification, and they could be paired with least-privilege auditor access.

References

[1] VentureBeat. "What the 'Blue Wave' in the 2026 U.S. Midterm Election Means for the American AI Industry and Businesses." VentureBeat, November 4, 2026.

[2] ABC7 News. "Midterm Election 2026 Live Results: Dems Take Back Congress." ABC7 News, November 4, 2026.

[3] Office of Rep. Lori Trahan. "Trahan, Obernolte Introduce Bipartisan FRONTIER Act to Strengthen Oversight of Advanced AI." U.S. House of Representatives, July 23, 2026.

[4] Statt. "The FRONTIER Act Explained: What H.R. 9925 Means for Frontier AI Regulation." Statt, July 27, 2026.

[5] Davis Wright Tremaine. "NY Overhauls Transparency and Governance Requirements for Frontier AI Developers." AI Law Advisor, April 2026.

[6] Hunton Andrews Kurth. "Illinois Governor Signs Frontier AI Model Law." Privacy and Cybersecurity Law Blog, July 2026.

[7] Miles Brundage. "Frontier AI Auditing-Related Legislation in the US: Landscape, Challenges, and a Path Forward." AVERI, April 20, 2026.

[8] Alston & Bird. "White House Secures Voluntary Industry Commitments on Independent AI Audits." Alston & Bird Privacy, Cyber & Data Strategy Blog, October 1, 2026.

[9] Campus Technology. "California Governor Proposes Independent Watchdogs Within AI Safety System." Campus Technology, September 23, 2026.

[10] Office of the Governor of California. "Executive Order N-9-26." State of California, September 18, 2026.

[11] DefenseScoop. "Bipartisan Senate Bill Would Push DOD to Expand Its Oversight of In-Use Commercial Frontier AI Models." DefenseScoop, October 8, 2026.

[12] The White House. "Fact Sheet: President Donald J. Trump Ensures a National Policy Framework for Artificial Intelligence." The White House, December 2025.

[13] California Legislature. "Business and Professions Code Section 22757.13." California Public Law, accessed November 2026.

[14] European Union. "EU AI Act, Article 73: Reporting of Serious Incidents." Regulation (EU) 2024/1689, 2024.

[15] Cybersecurity and Infrastructure Security Agency. "Cyber Incident Reporting for Critical Infrastructure Act of 2022 (CIRCIA)." CISA, accessed November 2026.

[16] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.1." CSA, 2026.

[17] Cloud Security Alliance. "AICM v1.1 Auditing Guidelines for Model Providers (MP)." CSA, 2026.

[18] Cloud Security Alliance. "Cloud Incident Response Framework – A Quick Guide." CSA, 2024.

[19] Cloud Security Alliance. "The Evolution of STAR: Introducing Continuous Auditing." CSA, 2024.

[20] European Union. "Regulation (EU) 2024/1689 (Artificial Intelligence Act)." Official Journal of the European Union, July 12, 2024.

[21] Lewis Silkin. "The Digital Omnibus on AI Enters into Force Today." Lewis Silkin, July 27, 2026.

[22] European Commission. "The General-Purpose AI Code of Practice." Shaping Europe's Digital Future, 2025.