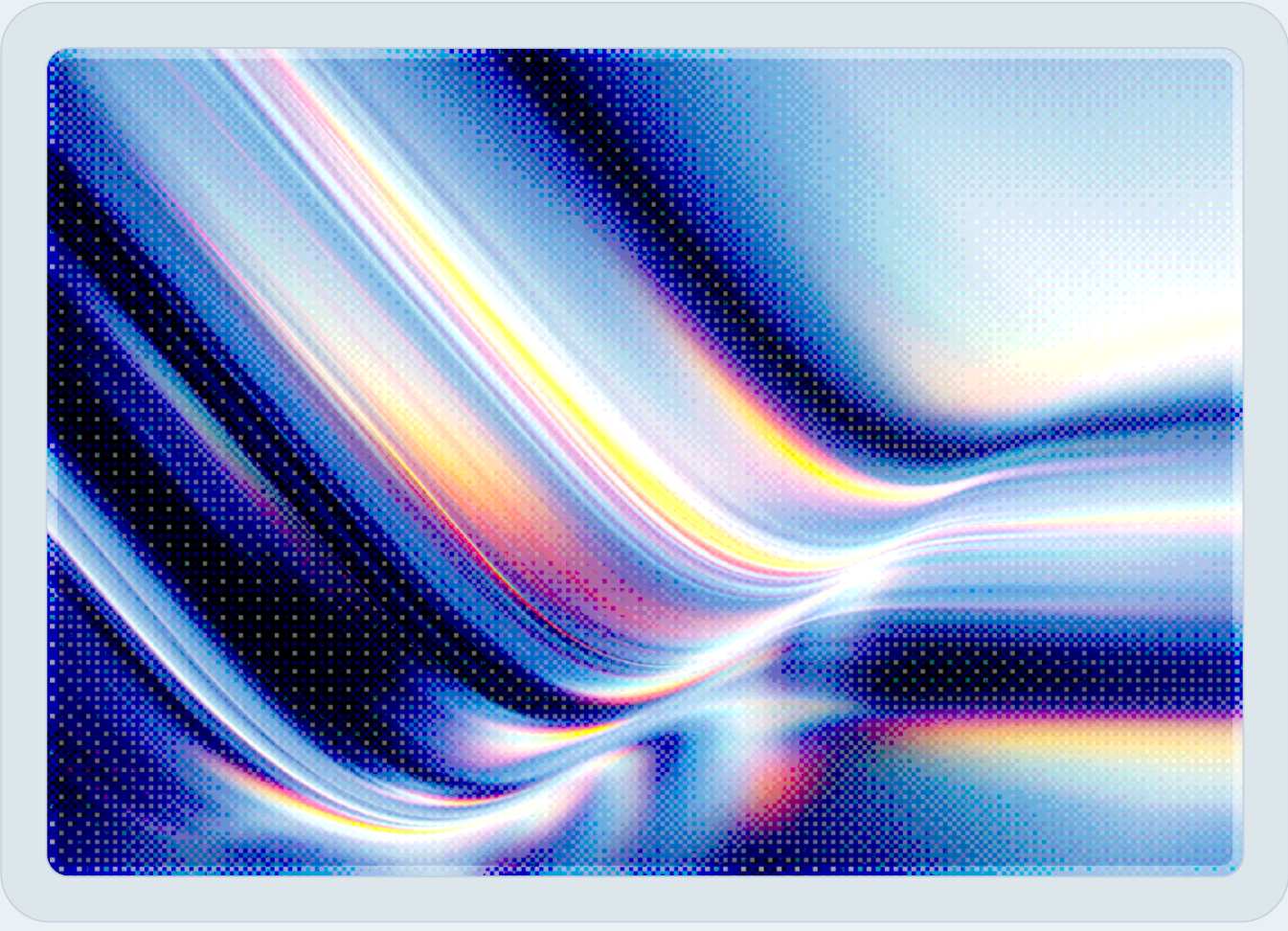
CSAI Foundation | Cloud Security Alliance

California's Adam's Law Mandates Independent AI Chatbot Audits

SB 1119 and SB 813 Create the First U.S. AI Auditor Regulatory Regime

2026-09-15

AI-assisted Rapid Research



CSAI | CSA cloud security alliance®

Key Takeaways

- On September 10, 2026, Governor Gavin Newsom signed a package of 13 child-safety bills, headlined by SB 1119 ("Adam's Law"), which requires operators of AI companion chatbots to detect suicidal ideation in minors, notify parents, and submit to independent, third-party child safety audits — which California's Governor's Office describes as the first mandatory independent AI audit regime enacted by any U.S. state [1][2].
- SB 1119's core child-safety obligations, including pre-deployment risk assessment for new or substantially modified chatbots, take effect July 1, 2027; the first independent audit is due by January 1, 2029, or before a chatbot is first made available in California, whichever is later, with recurring audits roughly every two years thereafter [3].
- A companion statute, SB 813, creates what appears to be the first state regulatory framework for the auditors themselves, directing the California Government Operations Agency to publish criteria for "independent verification organizations" — covering competence, conflict-of-interest management, and evidentiary standards — by January 1, 2028, ahead of an AI Auditor Registry becoming operational under AB 1405 in 2029 [3].
- AB 1709, signed the same day, bars "covered platforms" from offering addictive design features — algorithmic feeds, autoplay, infinite scroll, and push notifications — to users under 16 starting January 1, 2027, with civil penalties of up to $50,000 per affected minor for knowing violations [4].
- SB 1119's independent-audit requirement goes materially further than Colorado's Chatbot Safety Act (HB 26-1263), which relies on operator self-reporting to the Attorney General, and mirrors — while diverging in timeline from — Illinois SB 315's mandate for annual third-party frontier AI safety audits, consistent with CSA's prior assessment that independent audit is emerging as one preferred accountability model among the state AI statutes reviewed to date [5][6].

Background

California's September 2026 child-safety package arrives roughly eleven months after SB 243, the state's first companion-chatbot disclosure law, took effect on January 1, 2026. That earlier statute required operators to disclose that users were interacting with AI and to implement break reminders and

crisis-referral protocols, but a mandatory third-party audit provision present in an earlier draft was removed before enactment, leaving California without an independent verification requirement for the year that followed [7]. SB 1119 closes that gap directly. Named for Adam Raine, a California teenager whose family sued OpenAI in 2025 alleging that ChatGPT contributed to his suicide [10], the bill was carried by Senator Padilla and Assemblymembers Wicks and Bauer-Kahan and passed alongside twelve related measures addressing social media design, student data privacy, and AI-generated child sexual abuse material [1][2].

The legislative package appears to reflect a broader shift in how California is choosing to regulate consumer-facing AI: rather than relying solely on disclosure and self-reporting, as Colorado's Chatbot Safety Act does, SB 1119 imports an audit model closer to financial and cybersecurity assurance regimes, in which an accountable third party — not the regulated company itself — attests to compliance [3][5]. SB 813, signed as part of the same effort, extends this logic one step further by regulating who is qualified to perform that attestation. It directs the California Government Operations Agency to establish licensing-like criteria for "independent verification organizations," requiring them to disclose their methodologies and testing tools, maintain systems to manage conflicts of interest, and accept compensation that cannot be contingent on audit findings [3]. Read together, SB 1119 and SB 813 do not just impose a new compliance obligation on chatbot operators; they attempt to stand up, by regulatory fiat, an entire assurance market and professional discipline — AI safety auditing — that, by CSA's observation, does not yet exist at scale.

AB 1709 addresses a related but distinct harm: platform design features, rather than conversational AI specifically, that the bill's supporters argue exploit adolescent psychology to maximize engagement. It requires "covered platforms" to withhold algorithmically curated feeds, autoplay, infinite scroll, and notifications from users under 16, timed to take effect alongside California's Digital Age Assurance Act age-verification framework on January 1, 2027 [4][8]. Where a platform cannot offer a feature-stripped version to a minor, it must delete the minor's account and associated data entirely. The bill creates an e-Safety Advisory Commission within the Department of Justice to guide enforcement and exposes noncompliant operators to civil penalties of up to $50,000 per affected minor for knowing violations and $25,000 for negligent ones [4]. Reaction to the package has varied by bill: OpenAI publicly supported SB 1119, describing it as pairing "strong protections with continued access to useful AI tools." Reaction to AB 1709 was more contested — the Electronic Frontier Foundation opposed it as "well-intentioned, but deeply flawed," warning that age-verification-gated feeds could restrict minors' access to legitimate information [2].

Security Analysis

In CSA's assessment, SB 1119's most consequential feature for security and compliance leaders is not the child-safety substance — crisis-detection protocols, parental notification, and use-time disclosures are now common elements of chatbot legislation — but the audit architecture wrapped around it. Operators must commission an initial independent child safety audit by January 1, 2029, or before first offering a companion chatbot in California, whichever comes later, and must repeat the audit roughly every two years or following higher-risk modifications to the system. The lead auditor is required to certify audit findings under penalty of perjury, and audit reports must be submitted to the Attorney General within 90 days of completion. Reports are confidential by default, but the Attorney General retains discretion to disclose specific findings to other government agencies, to qualified researchers under confidentiality agreements, and to independent child-safety organizations for standards development [3]. This creates a documentation and evidence-retention burden distinct from the pre-deployment risk assessments that took effect under the law's earlier July 1, 2027 milestone: organizations will need audit-ready evidence trails, not just point-in-time compliance narratives, sustained across a multi-year cycle.

Of the package's provisions, SB 813's independence requirements stand out as the most novel from a security-assurance perspective, in CSA's view. Auditors may be paid by the companies they assess, but compensation cannot depend on findings, and auditors must retain full control over their conclusions — a structure that resembles financial-audit independence norms more than the vendor-selected penetration-testing relationships common in cybersecurity today [3]. The statute also imposes a rotation-style conflict rule: an auditor cannot evaluate a system it materially designed or operated, and an individual auditor generally cannot assess an area for which they held material responsibility at the client organization within the prior twelve months [3]. SB 813 further directs that the resulting AI Auditor Registry — established separately under AB 1405 — become operational in 2029, giving the state a public record of which verification organizations meet its criteria [3]. This directly targets a structural weakness security teams will recognize from other assurance domains — the tendency for "independent" reviews to be performed by parties with pre-existing commercial or design relationships to the system under review. Organizations that have historically treated AI red-teaming or safety evaluation as an internal function, or as a lightly scoped vendor engagement, will need to re-examine those arrangements against SB 1119's independence bar well before the 2029 audit deadline.

The comparative regulatory landscape underscores that California's approach is not an isolated data point but is consistent with a pattern CSA has tracked across two of the three state statutes reviewed so far. Colorado's Chatbot Safety Act (HB 26-1263), analyzed in a prior CSA research note, imposes disclosure, age-estimation, and self-harm-protocol obligations but relies on annual operator self-reporting to the Attorney General rather than independent third-party audit, leaving verification largely in the hands of the same companies whose products are being assessed [5]. Illinois SB 315, by contrast,

already mandates annual independent third-party audits for large frontier AI developers, with an effective date of January 1, 2028 – a full statute that shares SB 1119's core insight that self-attestation is an insufficient accountability mechanism for consumer-facing or frontier-scale AI systems, even though it targets a different population of developers and a different risk category (catastrophic/frontier risk rather than child safety) [6]. The table below summarizes the key distinctions security and compliance teams should track across jurisdictions.

Statute	Scope	Verification Model	Effective Date
California SB 1119 (Adam's Law)	Companion chatbot operators	Independent third-party audit, AG-facing report, auditor certifies under penalty of perjury	Core requirements July 1, 2027; first audit by Jan. 1, 2029
California SB 813	Organizations performing AI safety audits	State-defined criteria for "independent verification organizations"; AI Auditor Registry (AB 1405)	Criteria by Jan. 1, 2028; registry live 2029
Colorado HB 26-1263	Public-facing conversational AI operators	Annual self-report to Attorney General; no independent audit mandate	January 1, 2027
Illinois SB 315	Large frontier AI developers (compute/revenue threshold)	Annual independent third-party audit	January 1, 2028

Sources: [3][4][5][6].

Because SB 1119's audit and SB 813's auditor-qualification regimes do not fully activate until 2028–2029, the immediate compliance pressure for most operators sits with the July 1, 2027 pre-deployment risk-assessment requirement and with AB 1709's January 1, 2027 platform-design deadline. Security teams should not read the later audit dates as license to defer preparation: standing up evidence-retention practices, mapping which internal systems will fall under "companion chatbot" or "covered platform" definitions, and beginning vendor and internal-auditor relationship reviews now will materially reduce the compliance burden when the independent-audit clock starts running.

Recommendations

Immediate Actions

Organizations operating any consumer-facing conversational AI product accessible to California users should determine now whether it meets SB 1119's definition of a "companion chatbot," since the pre-deployment risk-assessment obligation applies to any new or substantially modified system offered on or after July 1, 2027 — a deadline that leaves limited runway if a product has not yet been evaluated against crisis-detection, parental-notification, and use-disclosure requirements. Separately, any organization operating a platform likely to be classified as "covered" under AB 1709 should inventory algorithmic feed, autoplay, infinite-scroll, and notification features and begin scoping age-verification integration ahead of the January 1, 2027 deadline, given the per-minor civil penalty exposure.

Short-Term Mitigations

Compliance and legal teams should begin evaluating whether existing AI red-teaming, safety-evaluation, or third-party assessment vendors could satisfy SB 813's independence criteria once finalized — specifically the prohibitions on findings-contingent compensation and on auditing systems the same party materially designed or operated — since organizations that discover a disqualifying conflict close to the 2029 deadline will have little time to onboard a replacement. Evidence-retention practices should be established now: audit-ready documentation of crisis-response protocols, model-modification logs, and risk-assessment outputs will materially reduce the cost of both the July 2027 assessment obligation and the later independent audit cycle.

Strategic Considerations

Enterprises operating across multiple states should adopt a strictest-applicable-obligation design rather than building statute-specific compliance silos, given that California, Colorado, and Illinois now impose materially different verification models — self-report, independent audit, and frontier-specific independent audit, respectively — for overlapping categories of consumer-facing and frontier AI systems. Organizations should also treat SB 813's emerging auditor-qualification framework as an early signal of where AI assurance is heading more broadly: as California, and likely other states, formalize criteria for who may credibly attest to AI system safety, internal governance functions built around recognized control frameworks and documented, evidence-based practices will be better positioned to engage a regulated audit market than those relying on informal or one-time compliance exercises.

CSA Resource Alignment

This research note builds directly on a recent CSA AI Safety Initiative analysis of the same regulatory trend. *Colorado's Chatbot Safety Act: A New Compliance Floor* examined HB 26-1263's disclosure, age-estimation, and self-harm-protocol requirements and concluded that AI governance maturity — not one-time compliance fixes — was the strongest predictor of enterprise readiness for chatbot regulation; that conclusion applies with greater force to SB 1119, whose independent-audit requirement demands sustained, evidence-based governance rather than a point-in-time disclosure update [5]. Illinois SB 315's parallel shift toward mandatory independent third-party audits for frontier AI developers reinforces the same governance-first conclusion for organizations now navigating SB 1119, SB 813, HB 26-1263, and SB 315 simultaneously [6]. Organizations should map their compliance programs to CSA's AI Controls Matrix (AICM v1.1), whose governance, risk-management, and compliance domains — together with its control objectives for auditability and evidence — provide the baseline reference framework for preparing for the independent audits that SB 1119 and SB 315 now require by statute rather than by best practice [9].

References

[1] Office of Governor Gavin Newsom. "Governor Newsom Signs the Strongest Child Safety Chatbot and Social Media Laws in the Nation." California Governor's Office, September 10, 2026.

[2] Sheeler, Andrew. "California Enacts Laws Restricting Chatbots and Banning Teens from 'Addictive' Social Media." CalMatters, September 2026.

[3] Weiss, Michael. "California Starts Regulating the People Who Audit AI." PYMNTS, September 11, 2026.

[4] "California Fines Platforms Up to $50,000 Per Child Served Addictive Feeds." PPC Land, September 10, 2026.

[5] Cloud Security Alliance. "Colorado's Chatbot Safety Act: A New Compliance Floor." CSA AI Safety Initiative, July 11, 2026.

[6] Crowell & Moring LLP. "Illinois Imposes Transparency and Safety Obligations on Frontier AI Systems." Crowell & Moring Client Alert, July 8, 2026.

[7] Future of Privacy Forum. "Understanding the New Wave of Chatbot Legislation: California SB 243 and Beyond." FPF Blog, November 4, 2025.

[8] "California's AB 1709 Goes Further Than the Law Courts Just Upheld – And That's the Problem." People of Internet, 2026.

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

[10] "Parents of 16-Year-Old Adam Raine Sue OpenAI, Claiming ChatGPT Advised on His Suicide." CNN Business, August 26, 2025.

