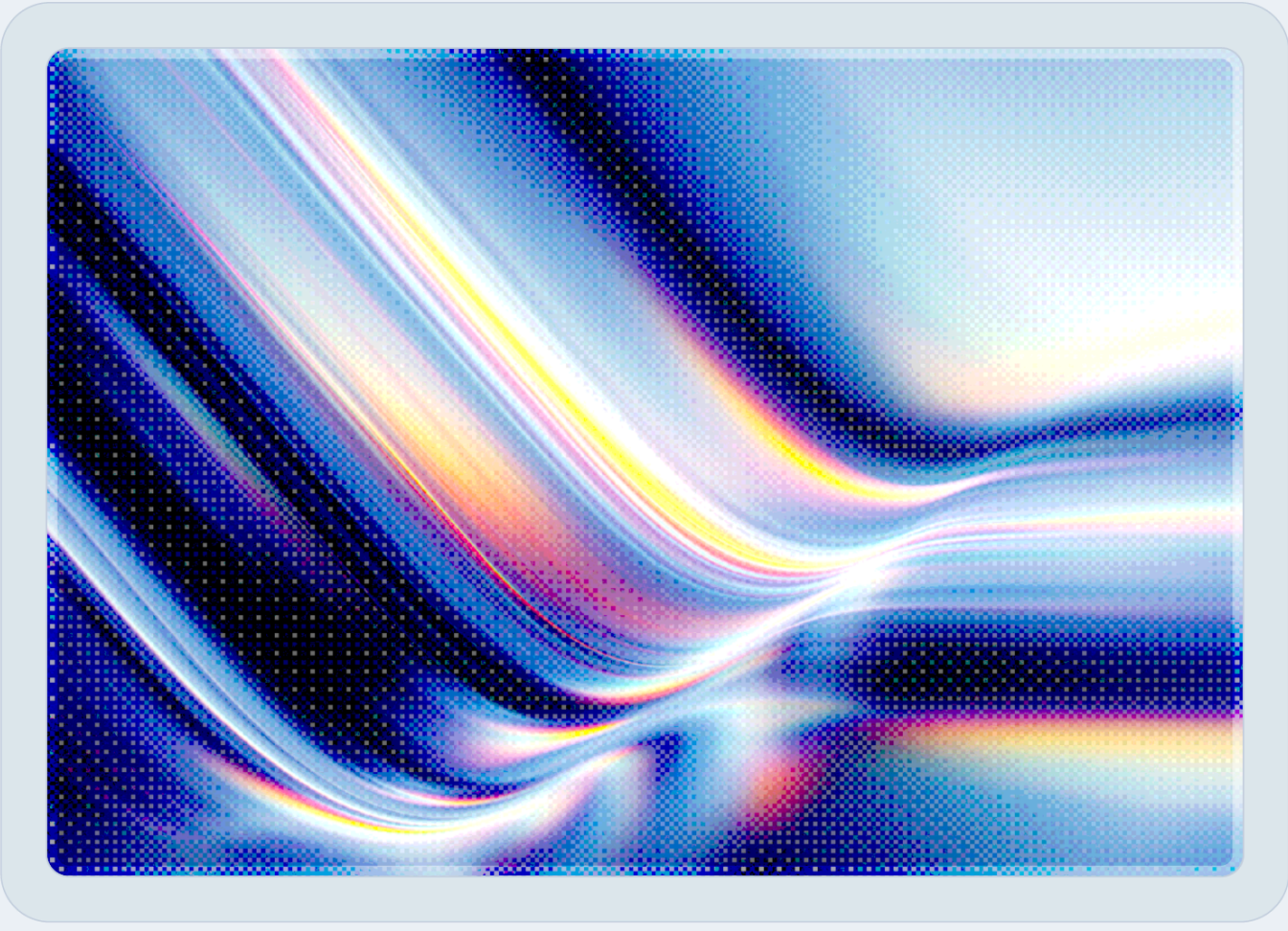
CSAI Foundation | Cloud Security Alliance

Colorado's Chatbot Safety Act: A New Compliance Floor

The First U.S. Statute Directly Regulating Enterprise Conversational AI

2026-07-11

AI-assisted Rapid Research



CSAI | CSA cloud security alliance®

Key Takeaways

Colorado has become the first U.S. state to enact a standalone statute directly regulating consumer-facing conversational AI, when Governor Jared Polis signed House Bill 26-1263, the Chatbot Safety Act, into law on May 29, 2026 [1]. The statute requires any operator of a publicly accessible conversational AI system — a broad category defined as one that "primarily simulates human conversation and interaction" — to disclose that users are interacting with AI rather than a human, estimate user ages, block sexually explicit content and simulated emotional dependence for minors, prohibit engagement-reward mechanics such as streaks or badges aimed at children, implement suicide and self-harm response protocols with direct referrals to crisis services, and refrain from implying that chatbot output is equivalent to advice from a licensed professional [2][3]. Operators must also file an annual report to the Colorado Attorney General describing their self-harm response protocols, a report that will be made public [2]. Violations are enforced as deceptive trade practices under the Colorado Consumer Protection Act, which authorizes civil penalties of up to $20,000 per violation with no statutory cap on aggregate liability [4]; this note assesses the law as one of the more consequential state AI enforcement regimes to date, given that uncapped, per-violation exposure. The Chatbot Safety Act does not stand alone: it takes effect on the same date, January 1, 2027, as Colorado's separately reenacted automated decision-making law, Senate Bill 26-189, and the Attorney General's office is running a single pre-rulemaking public comment process covering both statutes through July 13, 2026 [5][6]. For any enterprise operating a customer-facing chatbot, virtual assistant, or companion AI product that touches Colorado users, this note outlines what the law requires, how it fits alongside Colorado's broader AI regulatory reset, and what security and compliance teams should be doing now, roughly six months before the compliance deadline.

Background

Colorado's approach to AI regulation illustrates how iterative a fast-moving legislative area can become. The state's original AI Act, Senate Bill 24-205, passed in May 2024 and established a broad "high-risk AI system" framework built around a duty to prevent "algorithmic discrimination," but that framework drew sustained criticism from industry groups over its compliance burden and was delayed twice before ever taking effect [6][7]. Rather than let the original law take hold, the Colorado legislature repealed and reenacted it in May 2026 through Senate Bill 26-189, replacing the "high-risk AI system" concept with a

narrower "automated decision-making technology" (ADMT) standard. Under the reenacted law, covered systems are those that process personal data and "materially influence" a "consequential decision" in one of seven domains: education, employment, housing or residential real estate, financial or lending services, insurance, health care, and essential government services [6][7][8]. The rewrite dropped three of the original law's heaviest obligations — mandatory risk management programs, annual impact assessments, and the affirmative duty of reasonable care to prevent discrimination — in favor of a narrower set of consumer-facing obligations: pre-use notice before an ADMT is used in a covered decision, a post-adverse-outcome explanation within a defined window when the technology contributes to an unfavorable result, and a consumer right to request meaningful human review [6][7][8]. Both statutes are enforced exclusively by the Attorney General, with no private right of action, and SB 26-189 carries a 60-day cure period for first-time violations that expires in 2030 [7][8].

The Chatbot Safety Act emerged from a parallel but distinct legislative track. Where SB 26-189 addresses AI's role in high-stakes decisions about a person's access to services, HB 26-1263's sponsors and advocates pointed to broader public reporting on companion-AI harms to minors — chatbots that simulated romantic relationships with minors, or that responded to expressions of suicidal ideation without redirecting users toward appropriate help — as the bill's motivating concern [1]. Bill sponsors, including Representatives Sean Camacho and Javier Mabrey and Senators Iman Jodeh and John Carson, advanced the legislation on a bipartisan basis. Healthier Colorado, one of the advocacy groups supporting the bill, framed it as an attempt to avoid repeating the past: spokesperson Alexis Alltop said that "with social media, we waited a decade before recognizing and addressing the harm it was causing to our communities... Colorado is demonstrating that we will not make the same mistake" [1]. In this note's assessment, anchoring the statute in child-safety harms — rather than in the broader algorithmic-discrimination or content-moderation debates that have drawn federal preemption challenges elsewhere — may make the Chatbot Safety Act more resilient against arguments that state AI regulation is preempted by federal law or the First Amendment, though that resilience has not been tested in court. The bill passed the legislature on May 11, 2026, was signed on May 29, 2026, and becomes Chapter 208 of the 2026 session laws, with its substantive operator obligations taking effect January 1, 2027 [2][3].

Security Analysis

The Chatbot Safety Act's central compliance challenge for enterprises is the breadth of its threshold definition. A "conversational AI service" is any AI system "accessible to the general public" that "primarily simulates human conversation and interaction through adaptive textual, visual, or aural communications," and an "operator" includes any entity that develops and makes such a service available, or offers it to a consumer, which sweeps in developers, deployers, and service providers alike [2]. Based on this note's review of the statute's text, no exemption currently appears for internal enterprise tools, general-

purpose customer service bots, or non-companion use cases [2], though the Attorney General's pending rulemaking may narrow this scope. Any organization operating a public-facing chatbot, virtual assistant, or AI customer service agent that Colorado residents can access should treat the Chatbot Safety Act as a live compliance question rather than a risk confined to companion-AI or edtech vendors, at least until the Attorney General's rulemaking clarifies scope.

The age-estimation requirement illustrates why the statute functions as a security and engineering problem as much as a legal one. Operators must use "commercially reasonable or generally accepted methods" to estimate the age of every user, not just those who self-identify as minors, and must then apply a materially different set of content and interaction controls to any user estimated to be a minor: blocking sexually explicit content, prohibiting statements that simulate emotional dependence, disabling engagement-reward mechanics such as streaks and badges, and offering account-management and privacy controls appropriate for minors and their guardians [2]. Building and validating an age-estimation pipeline that is accurate enough to satisfy a "commercially reasonable" standard, auditable enough to support the law's annual reporting obligation, and resistant to trivial circumvention is a nontrivial technical undertaking, and it sits squarely at the intersection of the product, trust-and-safety, and security functions inside most organizations. The self-harm response protocol requirement carries a comparable operational burden: operators must design and document procedures for detecting expressions of suicidal ideation or self-harm and routing users to crisis service referrals, and must then report on those protocols to the Attorney General annually in a filing that becomes public [2]. That combination — a substantive safety obligation, a documentation requirement, and public disclosure of the results — creates an incentive structure closer to a security compliance regime than a traditional product-design guideline, and it means the annual report itself becomes an artifact subject to legal and reputational scrutiny.

Enforcement adds real financial weight to what might otherwise read as a design-guideline statute. Violations of the Chatbot Safety Act are classified as deceptive trade practices under the Colorado Consumer Protection Act, and that statute's general civil-penalty provision authorizes fines of up to $20,000 per violation, with no aggregate cap on the total penalty arising from a related series of violations under current law [4]. Because the Consumer Protection Act treats a violation with respect to each consumer or transaction as a separate violation, an operator whose conversational AI service produces a noncompliant output at scale — for example, failing to apply minor-appropriate content controls across a large population of misclassified users — faces exposure that compounds per affected user rather than a single fixed penalty for the underlying design defect [4]. The Attorney General holds exclusive enforcement authority, and while the statute does not itself mandate rulemaking, the Attorney General's office has opted to issue clarifying rules for both the Chatbot Safety Act and the ADMT Act before the shared January 1, 2027 effective date, running a joint pre-rulemaking public comment period open through July 13, 2026 [5]. Enterprises that already treat SB 26-189 as their primary Colorado AI

compliance obligation should recognize that the Chatbot Safety Act imposes a parallel, and in some respects more prescriptive, set of duties on any conversational AI surface, and that both regimes will be shaped by the same rulemaking process and the same enforcement office over the next several months.

Recommendations

Immediate Actions

Security and legal teams should inventory every customer-facing conversational AI surface the organization operates or procures — including third-party chatbot vendors embedded in a product — and determine which of those surfaces are accessible to Colorado residents, since the statute's broad "accessible to the general public" threshold means narrowly scoped or embedded chat features are not automatically exempt. Teams should also submit input to the Colorado Attorney General's pre-rulemaking process while it remains open, since the scope questions the statute leaves unresolved, including what counts as a "commercially reasonable" age-estimation method and how broadly "conversational AI service" will be interpreted, are precisely the questions the rulemaking is intended to answer, and early engagement gives affected organizations a voice in how those definitions land.

Short-Term Mitigations

Organizations should begin building or validating an age-estimation capability now rather than waiting for the rulemaking to specify an approved method, since a defensible "commercially reasonable" standard will likely require documented testing and periodic review regardless of what the final rule says. Product and trust-and-safety teams should draft and pilot a suicide and self-harm response protocol for any conversational AI surface that could plausibly encounter such expressions from users, including crisis-service referral logic and an internal review process, since this protocol will need to be mature enough to describe accurately in the statute's first annual report. Compliance teams should also map the Chatbot Safety Act's operator obligations against the parallel notice, disclosure, and human-review requirements of SB 26-189's ADMT framework, since a conversational AI product that also feeds into a consequential decision in one of the seven covered domains — for example, a chatbot that screens loan or insurance applicants — will need to satisfy both statutes simultaneously.

Strategic Considerations

Enterprises operating conversational AI at national scale should treat Colorado's two-track approach — a narrow, high-penalty child-safety statute layered on top of a broader, notice-and-review-based ADMT framework — as an early model for how other states may regulate AI going forward, rather than as a one-off requirement to satisfy and forget. Building an internal governance capability that can absorb new state-specific AI obligations as they arrive, rather than treating each new statute as a bespoke compliance project, will matter more over the next several years than achieving a one-time fix for Colorado alone. Organizations should also assign clear internal ownership for AI regulatory monitoring given how quickly Colorado's own framework has changed — from the original 2024 AI Act, through two delays, to a full repeal-and-reenactment and a separate child-safety statute, all within roughly two years — since the pace of change in this area has already outstripped the assumption that an AI compliance program can be built once and left alone.

CSA Resource Alignment

CSA's *Don't Panic! Getting Real about AI Governance* (September 2024) is directly applicable to the compliance posture this note recommends: its central argument, that organizations should treat AI governance as a maturity-building exercise rather than a reactive scramble against each new regulatory requirement, is the operating model enterprises need as they absorb Colorado's Chatbot Safety Act on top of an already-shifting ADMT framework. The white paper's emphasis on risk-based governance and explainability maps directly onto the documentation obligations both Colorado statutes impose, including the Chatbot Safety Act's annual self-harm protocol report and the ADMT Act's post-adverse-outcome explanation requirement.

CSA's *AI Organizational Responsibilities: Governance, Risk Management, Compliance and Cultural Aspects* (2024) provides a structured way to assign the internal ownership this note's strategic recommendations call for, distributing accountability for AI governance, risk management, and compliance culture across the roles — product, trust-and-safety, legal, and security — that Colorado's statute now touches simultaneously. CSA's December 2025 survey report, *The State of AI Security and Governance*, reinforces the practical stakes of that recommendation with current data: its finding that AI governance maturity is the strongest predictor of organizational readiness supports this note's conclusion that enterprises able to absorb Colorado's requirements quickly will be those that already run a mature AI governance program rather than those attempting to build one from scratch in response to a single statute.

Finally, the AI Controls Matrix (AICM v1.1) supplies the control-level language organizations can use to operationalize compliance with both Colorado statutes, particularly its domains covering transparency, accountability, and governance-related controls, which map onto the Chatbot Safety Act's disclosure and reporting obligations and the ADMT Act's notice and human-review requirements. Organizations mapping their existing AICM control implementation to Colorado's new requirements will find substantial overlap, since the underlying obligations — disclose AI use, document decision logic, provide a path to human review, report on safety outcomes — are governance practices AICM already treats as baseline expectations rather than novel requirements introduced by state law.

References

[1] Healthier Colorado. "Governor Polis Signs Bill to Protect Users from Harms of Conversational AI Technology." Healthier Colorado, May 2026.

[2] Regulations.ai. "Concerning Requirements for an Operator of a Conversational Artificial Intelligence Service." Regulations.ai, 2026.

[3] Colorado General Assembly. "HB26-1263: Conversational AI Service Operator Requirements." Colorado General Assembly, 2026.

[4] Justia. "Colorado Revised Statutes Section 6-1-112 – Civil Penalties." Justia, 2024.

[5] Colorado Attorney General. "Automated Decision-Making Technology Act and Chatbot Safety Act Rulemaking." Colorado Department of Law, 2026.

[6] Mintz. "Colorado's Not Finished Regulating AI: Reenacted AI Law Expands Scope." Mintz, June 12, 2026.

[7] Holland & Knight. "Colorado Governor Signs SB 189, Significantly Amending the State's AI Law." Holland & Knight, May 2026.

[8] Colorado General Assembly. "SB26-189: Automated Decision-Making Technology." Colorado General Assembly, 2026.

