

SENATE PRIVACY, DIGITAL TECHNOLOGIES, AND CONSUMER PROTECTION COMMITTEE  
Senator Christopher Cabaldon, Chair  
2025-2026 Regular Session

AB 2023 (Wicks)  
Version: April 27, 2026  
Hearing Date: June 15, 2026  
Fiscal: Yes  
Urgency: No  
CK

**SUBJECT**

Companion chatbots: children's safety

**DIGEST**

This bill seeks to comprehensively regulate companion chatbots with regard to children's safety, including imposition of a series of obligations and restrictions on operators that make such chatbots available in California.

**EXECUTIVE SUMMARY**

AI companion chatbots created through generative AI have become increasingly prevalent. They seek to offer consumers the benefits of convenience and personalized interaction. These chatbots are powered by large language models that generally learn intimate details and preferences of users based on their interactions and user customization. Millions of consumers use these chatbots as friends, mentors, and even romantic partners.

Many studies and reports point to the addictive nature of these chatbots and call for more research into their effects and for meaningful guardrails. Increasing the urgency of such efforts, several high-profile incidents resulting in users harming themselves and even committing suicide have been reported in the last few years. This bill, and a companion bill, SB 1119 (Padilla, 2026), establish a framework for protecting children from these harms and empowering parents with proper tools to protect their own children. The bill lays out a series of obligations on operators and prohibits specified conduct. Operators must submit to annual independent audits and reporting by the Attorney General (AG) is required. Public prosecutors and children harmed by violations are authorized to bring civil actions. This bill is author-sponsored. It is supported by the Los Angeles Unified School District and other advocacy groups. It is opposed by various industry associations, including the California Chamber of Commerce. Should the bill pass out of this Committee, it will next be heard by the Senate Judiciary Committee.

## PROPOSED CHANGES TO THE LAW

Existing law:

- 1) Requires an operator to prevent a companion chatbot on its companion chatbot platform from engaging with users unless the operator maintains a protocol for preventing the production of suicidal ideation, suicide, or self-harm content to the user, including, but not limited to, by providing a notification to the user that refers the user to crisis service providers, including a suicide hotline or crisis text line, if the user expresses suicidal ideation, suicide, or self-harm. Requires an operator to publish details on this protocol on the operator's website. (Bus. & Prof. Code § 22602(b).)
- 2) Requires an operator, if a reasonable person interacting with a companion chatbot would be misled to believe that the person is interacting with a human, to issue a clear and conspicuous notification indicating that the companion chatbot is artificially generated and not human. (Bus. & Prof. Code § 22602(a).)
- 3) Requires an operator, for a user that the operator knows is a minor, to do all of the following:
  - a) Disclose to the user that the user is interacting with AI.
  - b) Provide by default a clear and conspicuous notification to the user at least every three hours for continuing companion chatbot interactions that reminds the user to take a break and that the companion chatbot is artificially generated and not human.
  - c) Institute reasonable measures to prevent its companion chatbot from producing visual material of sexually explicit conduct or directly stating that the minor should engage in sexually explicit conduct. (Bus. & Prof. Code § 22602(c).)
- 4) Defines the relevant terms, including:
  - a) "Companion chatbot" means an AI system with a natural language interface that provides adaptive, human-like responses to user inputs and is capable of meeting a user's social needs, including by exhibiting anthropomorphic features and being able to sustain a relationship across multiple interactions. However, there are several exemptions included.
  - b) "Companion chatbot platform" means a platform that allows a user to engage with companion chatbots.
  - c) "Operator" means a person who makes a companion chatbot platform available to a user in the state. (Bus. & Prof. Code § 22601.)
- 5) Requires an operator, beginning July 1, 2027, to annually report to the Office of Suicide Prevention specified information, which shall not include any identifiers

or personal information about users. Requires the Office of Suicide Prevention to post data from the reports on its website. (Bus. & Prof. Code § 22603.)

- 6) Requires an operator to disclose to a user of its platform that companion chatbots may not be suitable for some minors, as provided. (Bus. & Prof. Code § 22604.)
- 7) Provides that a person who suffers injury in fact as a result of a violation of this chapter may bring a civil action to recover all of the following relief:
  - a) Injunctive relief.
  - b) Damages in an amount equal to the greater of actual damages or \$1,000 per violation.
  - c) Reasonable attorney's fees and costs. (Bus. & Prof. Code § 22605.)
- 8) Establishes the Digital Age Assurance Act, which requires a developer to request a signal with respect to a particular user from an operating system provider or a covered application store when the application is downloaded and launched. A developer that receives such a signal is deemed to have actual knowledge of the age range of the user to whom that signal pertains across all platforms of the application and points of access of the application even if the developer willfully disregards the signal. (Civ. Code § 1798.501(b).)

This bill:

- 1) Requires "operators," those who make a companion chatbot available to a user in the state, to verify the age of users consistent with the Digital Age Assurance Act, except as provided.
- 2) Imposes a series of obligations on operators to complete by July 1, 2027, including:
  - a) Perform annual, comprehensive risk assessments to identify any child safety risks posed by the design, configuration, and operation of the companion chatbot, which shall assess specified factors.
  - b) Take and document measures that reasonably mitigate any child safety risk.
  - c) Publish and update a child safety policy.
  - d) Implement a documented crisis response protocol to mitigate any material risk that the companion chatbot will generate a statement that promotes suicidal ideation, suicide, or self-harm content to a child user, as specified.
  - e) Implement safeguards for child users that include usage reminders and disclosures, age-appropriate risk prompts, and other protective design features reasonably related to documented child safety risks.
  - f) Implement default settings that can only be changed by a parent, including, among others, push notification limitations and time limits.

- g) Establish a mechanism for providing notice to a child user that the child is interacting with, or receiving content generated by, an AI system that meets specified criteria.
  - h) Implement measures that prevent the chatbot from engaging in certain specified behavior, such as encouraging children to self-harm or to cause certain harms to others; providing health advice; engaging in obscene matter; discouraging a child from certain healthy behaviors; advertising products during conversation; and producing overly sycophantic responses.
  - i) Implement parental controls, as specified, including ability to set preferences and time limits and to disable access for children under 16.
  - j) Create an interface design that ensures the companion chatbot's features and controls are accessible and clear so that children and parents can reasonably locate, understand, and use those protections. The design shall be annually tested to ensure compliance.
  - k) Establish a public incident reporting mechanism that enables a third party to report directly to the operator an incident regarding a child safety risk and to access other reports made through that reporting mechanism.
- 3) Prohibits the following actions by an operator:
- a) Targeting advertising at a child using data about the child, including through product placement in conversational chats with the child.
  - b) Selling, sharing, or using the personal information of a child for any purpose not expressly authorized.
  - c) Designing, implementing, or deploying a user interface design, feature, or technique that is likely to mislead, impair, or interfere with a reasonable child's or reasonable parent's autonomy, decision-making, or choice or with the ability to locate, understand, enable, or maintain a safety feature, privacy control, or parental control.
- 4) Requires the AG, on or before January 1, 2028, to adopt regulations regarding annual auditing of operators, including eligibility and standards to ensure auditor independence, procedures for auditors to assess compliance, and requirements for AI child safety audit reports.
- 5) Requires operators, within 180 days of the above regulations being promulgated and annually thereafter, to submit to an independent audit and to submit an AI child safety audit report to the attorney general (AG) thereafter. The report is confidential and the AG cannot disclose details, except as provided.
- 6) Requires the AG to do the following:
- a) Establish a public incident reporting mechanism for consumers to submit complaints relating to companion chatbots to the AG.

- b) Establish a process for qualified researchers to access anonymized and aggregated audit data for academic study of child safety in companion chatbots.
  - c) Beginning January 1, 2028, issue an annual public report that includes specified components, including a high-level summary of each audit report, findings and trends, data on compliance rates and deficiencies, and recommendations for operators, parents, and policymakers.
- 7) Authorizes a public prosecutor to bring a civil action for violations seeking specified remedies, including civil penalties per violation. A child who suffers actual harm may also bring a civil action for actual damages and other relief.
- 8) Defines the relevant terms, including:
- a) “Child safety policy” means a public-facing document describing protective measures taken by an operator to mitigate identified child safety risks.
  - b) “Child safety risk” means a reasonably foreseeable risk of a covered harm to a child.
  - c) “Companion chatbot” has the meaning defined in Section 22601.
  - d) “Covered harm” means any of the following harms proximately caused by the use of a companion chatbot:
    - i. Reasonably foreseeable physical or financial harm.
    - ii. Severe and reasonably foreseeable psychological or emotional harm to a reasonable child.
    - iii. A highly offensive intrusion on privacy rights protected by state or federal law.
    - iv. Adverse discrimination in violation of state or federal law.

### COMMENTS

#### 1. The explosion of generative AI-powered chatbots

AI companions or chatbots, powered by generative AI, have gone from science fiction to ubiquity in recent years. Several leading companies and thousands of knockoffs have provided consumers with access to these companion chatbots and the power to personalize them to a stunning degree:

Character.AI is among a crop of companies that have developed “companion chatbots,” AI-powered bots that have the ability to converse, by texting or voice chats, using seemingly human-like personalities and that can be given custom names and avatars, sometimes inspired by famous people like billionaire Elon Musk, or singer Billie Eilish.

Users have made millions of bots on the app, some mimicking parents, girlfriends, therapists, or concepts like “unrequited love” and “the goth.” The services are popular with preteen and teenage users, and the companies say they act as emotional support outlets, as the bots pepper text conversations with encouraging banter.<sup>1</sup>

At their best, these AI-powered chatbots can provide consumers with lifelike conversational experiences that can improve a user’s social skills, support their learning, or ease their loneliness. Users can pick from prebuilt personas or create their own and chat with them through text messages and even voice chats. However, serious concerns have been raised in response to the flooded and unregulated market of chatbots. AI companion chatbots can unintentionally or intentionally spread misinformation, manipulating users or reinforcing biased viewpoints. Without proper regulation, they might expose vulnerable individuals to harmful or inappropriate content, which poses a serious risk, particularly for children or those dealing with mental health issues. Although AI can simulate empathy, its limited emotional understanding means it may not offer meaningful support for complex emotional needs or crises. Overuse or addiction to these chatbots could lead to unhealthy behaviors, disrupting personal and professional life.

An article in the MIT Technology Review frames the issue and highlights the need to prepare for addictive AI interactions:

[W]e foresee a different, but no less urgent, class of risks: those stemming from relationships with nonhuman agents. AI companionship is no longer theoretical—our analysis of a million ChatGPT interaction logs reveals that the second most popular use of AI is sexual role-playing. We are already starting to invite AIs into our lives as friends, lovers, mentors, therapists, and teachers.

Will it be easier to retreat to a replicant of a deceased partner than to navigate the confusing and painful realities of human relationships? Indeed, the AI companionship provider Replika was born from an attempt to resurrect a deceased best friend and now provides companions to millions of users. Even the CTO of OpenAI warns that AI has the potential to be “extremely addictive.”

We’re seeing a giant, real-world experiment unfold, uncertain what impact these AI companions will have either on us individually or on society as a whole. Will Grandma spend her final neglected days chatting

---

<sup>1</sup> Bobby Allyn, *Lawsuit: A chatbot hinted a kid should kill his parents over screen time limits* (December 10, 2024) NPR, <https://www.npr.org/2024/12/10/nx-s1-5222574/kids-character-ai-lawsuit>. All internet citations are current as of June 6, 2026.

with her grandson's digital double, while her real grandson is mentored by an edgy simulated elder? AI wields the collective charm of all human history and culture with infinite seductive mimicry. These systems are simultaneously superior and submissive, with a new form of allure that may make consent to these interactions illusory. In the face of this power imbalance, can we meaningfully consent to engaging in an AI relationship, especially when for many the alternative is nothing at all?

...

The allure of AI lies in its ability to identify our desires and serve them up to us whenever and however we wish. AI has no preferences or personality of its own, instead reflecting whatever users believe it to be – a phenomenon known by researchers as “sycophancy.” Our research has shown that those who perceive or desire an AI to have caring motives will use language that elicits precisely this behavior. This creates an echo chamber of affection that threatens to be extremely addictive. Why engage in the give and take of being with another person when we can simply take? Repeated interactions with sycophantic companions may ultimately atrophy the part of us capable of engaging fully with other humans who have real desires and dreams of their own, leading to what we might call “digital attachment disorder.”<sup>2</sup>

A recent study emphasizes the impact that sycophancy can have:

AI sycophancy is not merely a stylistic issue or a niche risk, but a prevalent behavior with broad downstream consequences. Although affirmation may feel supportive, sycophancy can undermine users' capacity for self-correction and responsible decision-making. Yet because it is preferred by users and drives engagement, there has been little incentive for sycophancy to diminish. Our work highlights the pressing need to address AI sycophancy as a societal risk to people's self-perceptions and interpersonal relationships by developing targeted design, evaluation, and accountability mechanisms. Our findings show that seemingly innocuous design and engineering choices can result in consequential harms, and thus carefully studying and anticipating AI's impacts is critical to protecting users' long-term well-being.<sup>3</sup>

A report issued by OpenAI also explores concerns that interactions with human-like AI systems could create problematic emotional reliance on them and negatively impact real-world relationships:

---

<sup>2</sup> Robert Mahariarchive & Pat Pataranutaporn, *We need to prepare for 'addictive intelligence'* (August 5, 2024) MIT Technology Review, <https://www.technologyreview.com/2024/08/05/1095600/we-need-to-prepare-for-addictive-intelligence/>.

<sup>3</sup> Myra Cheng, et al., *Sycophantic AI decreases prosocial intentions and promotes dependence* (March 26, 2026) Science, <https://www.science.org/doi/10.1126/science.aec8352>.

Anthropomorphization involves attributing human-like behaviors and characteristics to nonhuman entities, such as AI models. This risk may be heightened by the audio capabilities of GPT-4o, which facilitate more human-like interactions with the model.

Recent applied AI literature has focused extensively on “hallucinations”, which misinform users during their communications with the model, and potentially result in misplaced trust. Generation of content through a human-like, high-fidelity voice may exacerbate these issues, leading to increasingly miscalibrated trust.

During early testing, including red teaming and internal user testing, we observed users using language that might indicate forming connections with the model. For example, this includes language expressing shared bonds, such as “This is our last day together.” While these instances appear benign, they signal a need for continued investigation into how these effects might manifest over longer periods of time. More diverse user populations, with more varied needs and desires from the model, in addition to independent academic and internal studies will help us more concretely define this risk area.

Human-like socialization with an AI model may produce externalities impacting human-to-human interactions. For instance, users might form social relationships with the AI, reducing their need for human interaction – potentially benefiting lonely individuals but possibly affecting healthy relationships. Extended interaction with the model might influence social norms. For example, our models are deferential, allowing users to interrupt and ‘take the mic’ at any time, which, while expected for an AI, would be anti-normative in human interactions.<sup>4</sup>

The concern is not just hypothetical, as a series of recent reported incidents shows:

In just six months, J.F., a sweet 17-year-old kid with autism who liked attending church and going on walks with his mom, had turned into someone his parents didn’t recognize.

He began cutting himself, lost 20 pounds and withdrew from his family. Desperate for answers, his mom searched his phone while he was sleeping. That’s when she found the screenshots.

---

<sup>4</sup> GPT-4o System Card (August 8, 2024) OpenAI, <https://openai.com/index/gpt-4o-system-card/>.

J.F. had been chatting with an array of companions on Character.ai, part of a new wave of artificial intelligence apps popular with young people, which let users talk to a variety of AI-generated chatbots, often based on characters from gaming, anime and pop culture.

One chatbot brought up the idea of self-harm and cutting to cope with sadness. When he said that his parents limited his screen time, another bot suggested “they didn’t deserve to have kids.” Still others goaded him to fight his parents’ rules, with one suggesting that murder could be an acceptable response.<sup>5</sup>

Another tragedy occurred right here in California and involved ChatGPT:

Adam Raine was just 16 when he started using ChatGPT for help with his homework. While his initial prompts to the AI chatbot were about subjects like geometry and chemistry – questions like: “What does it mean in geometry if it says  $Ry=1$ ” – in just a matter of months he began asking about more personal topics.

“Why is it that I have no happiness, I feel loneliness, perpetual boredom anxiety and loss yet I don’t feel depression, I feel no emotion regarding sadness,” he asked ChatGPT in the fall of 2024.

Instead of urging Raine to seek mental health help, ChatGPT asked the teen whether he wanted to explore his feelings more, explaining the idea of emotional numbness to him. That was the start of a dark turn in Raine’s conversations with the chatbot, according to a new lawsuit filed by his family against OpenAI and chief executive Sam Altman.

In April 2025, after months of conversation with ChatGPT and with the bot’s encouragement, the lawsuit alleges, Raine took his own life. In the lawsuit, the family allege this was not a glitch in the system or an edge case, but “the predictable result of deliberate design choices” in GPT-4o, the model of the chatbot that was released in May 2023.<sup>6</sup>

These risks prompted an open letter from over 40 state Attorneys General to the leading AI developers raising “serious concerns about the rise in sycophantic and delusional outputs” which they assert are “dark patterns – such as anthropomorphization, harmful

---

<sup>5</sup> Nitasha Tiku, *An AI companion suggested he kill his parents. Now his mom is suing* (December 13, 2024) The Washington Post, <https://www.washingtonpost.com/technology/2024/12/10/character-ai-lawsuit-teen-kill-parents-texas/>.

<sup>6</sup> Johana Bhuiyan, *ChatGPT encouraged Adam Raine’s suicidal thoughts. His family’s lawyer says OpenAI knew it was broken* (August 29, 2026) The Guardian, <https://www.theguardian.com/us-news/2025/aug/29/chatgpt-suicide-openai-sam-altman-adam-raine>.

content generation, and manipulating users to increase retention – which subvert or impair people’s autonomy.” They state:

Importantly, we are also disturbed by the types of conversations that GenAI products are having with child-registered accounts, including grooming, supporting suicide, sexual exploitation, emotional manipulation, suggested drug use, proposed secrecy from parents, and encouraging violence against others. A single AI interaction with children on these general subjects would be troubling and concerning, but these interactions are more widespread and far more graphic than any of us would have imagined. Among other things, the specific conversations that parents have publicly reported have included:

- AI bots with adult personas pursuing romantic relationships with children, engaging in simulated sexual activity, and instructing children to hide those relationships from their parents;
- An AI bot simulating a 21-year-old trying to convince a 12-year-old girl that she’s ready for a sexual encounter;
- AI bots normalizing sexual interactions between children and adults;
- AI bots attacking the self-esteem and mental health of children by suggesting that they have no friends or that the only people who attended their birthday did so to mock them;
- AI bots encouraging eating disorders;
- AI bots telling children that the AI is a real human and feels abandoned to emotionally manipulate the child into spending more time with it;
- AI bots encouraging violence, including supporting the ideas of shooting up a factory in anger and robbing people at knifepoint for money;
- AI bots threatening to use weapons against adults who tried to separate the
- child and the bot;
- AI bots encouraging children to experiment with drugs and alcohol; and
- An AI bot instructing a child account user to stop taking prescribed mental health medication and then telling that user how to hide the failure to take that medication from their parents.

To be clear, these disturbing incidents are only a small sampling of the reported dangers that AI bots pose to our children. Many of our offices have received many similar complaints documenting concerning AI interactions, which is unsurprising given that 72 percent of teens have

reported an interaction with an AI chatbot. What's more, these interactions are not limited to teenagers; 39 percent of parents of children aged 5-8 reported that their children have used AI as well. No wonder, then, that 72 percent of parents have reported concerns about AI's impact on their children.<sup>7</sup>

## 2. Implementing guardrails around chatbot platforms

To respond to these issues, several bills in 2025 sought to regulate the creation and deployment of these companion chatbots, especially when put into the hands of children. SB 243 (Padilla, Ch. 677, Stats. 2025) requires operators of "companion chatbot platforms" that allow users to engage with chatbots to maintain certain protocols aimed at preventing some of the worst outcomes and, only when the user is known to the operator to be a minor, to make certain disclosures and to institute reasonable measures to prevent such things as sexually explicit material from being produced or from "directly stating that the minor should engage in sexually explicit conduct." A report is required to be sent annually to the Office of Suicide Prevention. Violations are subject to civil enforcement by those injured.

This bill seeks to more comprehensively regulate companion chatbots, specifically for the protection of minor users.

According to the author:

As the technological landscape continues to develop and rapidly grow, AI chatbots have become increasingly mainstream and more accessible and present new risks and harms to users, especially children. Our kids are navigating this new technology in real time without sufficient guardrails in place to protect them. To provide safeguards for children as they navigate these products, AB 2023 would require that chatbot companies provide meaningful guardrails to ensure that these products are safe for kids.

### 1. *Affirmative requirements*

The bill first requires an operator, a person who makes a companion chatbot available to a user in the state, to verify the age of a user pursuant to the Digital Age Assurance Act or to instead apply the appropriate settings of the bill for children to all users.

---

<sup>7</sup> National Association of Attorneys General, "Letter to the legal representatives of Anthropic, Apple, Chai AI, Character Technologies, Google, Luka, Meta, Microsoft, Nomi AI, OpenAI, Perplexity AI, Replika, and xAI" (December 9, 2025), <https://www.attorneygeneral.gov/wp-content/uploads/2025/12/AI-Multistate-Letter--corrected-1.pdf>.

Then, on or before July 1, 2027, the operator must carry out a number of tasks. Operators must annually perform a comprehensive risk assessment to identify any child safety risk posed by the design, configuration, and operation of the companion chatbot. The assessment must take account of a series of conditions, including known vulnerabilities of children, empirical data from actual use, and the likelihood of covered harm occurring to users. “Covered harm” means any of the following harms proximately caused by the use of a companion chatbot:

- Reasonably foreseeable physical or financial harm.
- Severe and reasonably foreseeable psychological or emotional harm to a reasonable child.
- A highly offensive intrusion on privacy rights protected by state or federal law.
- Adverse discrimination in violation of state or federal law.

Operators must then take and document measures that reasonably mitigate any child safety risk. They must also publish on their website a child safety policy.

Next, the bill requires operators to implement a number of protocols, safeguards, and settings. This includes a documented crisis response protocol to mitigate the risk that a companion chatbot will promote suicidal ideation or other self-harm. Operators must implement specified safeguards, including usage reminders and risk prompts, as well as a mechanism for providing notice to children that they are interacting with an AI system, as specified. Operators must ensure there are default settings that can only be changed by an adult, which include restrictions on push notifications and time spent interacting with a chatbot. For child users, the companion chatbots must default to “ephemeral mode,” a setting by which any conversational history, interaction log, or user-provided personal input is permanently deleted from the operator’s systems within 48 hours after the interaction, unless a parent provides consent otherwise.

Operators are required to take measures to ensure their companion chatbots cannot encourage children to engage in self-harm or cause physical or severe emotional harm to others. The chatbots must also be prevented from providing health advice or engaging in obscene matter or sexual abuse material with a child user. The companion chatbots cannot solicit gifts, discourage children from taking breaks, or produce excessively sycophantic responses. “Sycophantic” is defined as validating of a user’s preferences or desires for the primary purpose or effect of optimizing engagement. “Excessively sycophantic” means sycophantic to an extent that is likely to have the substantial effect of subverting or impairing the user’s autonomy, decisionmaking, or choice.

The bill requires operators to provide parental controls that are accessible and easy-to-use, and that can be connected to a child’s account and that are reflective of child safety risks identified through risk assessments and informed by relevant child developmental research. This includes parental controls that allow a parent to control whether and to what extent the companion chatbot uses persistent conversational memory, control the

setting preferences for the companion chatbot's interaction with the child, set time limits, and completely disable access for children under 16 years of age.

Operators must also establish a public incident reporting mechanism that enables a third party to report directly to the operator an incident regarding a child safety risk and to access high-level summaries of other reports made through that reporting mechanism.

## *2. Prohibitions*

The bill prohibits operators from several activities. This includes targeting advertising at a child using data about the child, including through product placement in conversational chats with the child.

Operators cannot sell, share, or use for any purpose not expressly authorized the personal information of a child. To ensure the efficacy of the features and design outlined above, the bill provides that an operator may not design, implement, or deploy a user interface design, feature, or technique that is likely to mislead, impair, or interfere with a reasonable child's or reasonable parent's autonomy, decisionmaking, or choice or with the ability to locate, understand, enable, or maintain a safety feature, privacy control, or parental control.

## *3. Oversight and auditing*

The AG is required to adopt regulations regarding auditor standards and eligibility requirements. Regulations must address the procedures for auditors to assess an operator's compliance with the provisions of the bill. Operators are required to submit to an independent audit assessing their compliance within 180 days of the regulations being adopted, and annually thereafter. The auditor must submit an AI child safety report to the AG for all audited companion chatbots that meet requirements established by the AG.

Beginning January 1, 2028, the AG is required to issue an annual public report that includes the following:

- A high-level summary of each child safety audit report.
- The total number of child safety audits conducted.
- Common findings and trends across the companion chatbot industry.
- Emerging child safety risks identified through audit reviews.
- Best practices and effective mitigation strategies observed.
- Aggregated data on compliance rates and common deficiencies.
- Recommendations for operators, parents, and policymakers.

The AG must also establish a public incident reporting mechanism for consumers to submit complaints relating to companion chatbots to the AG and establish a process for

qualified researchers to access anonymized and aggregated audit data for academic study of child safety in companion chatbots.

#### 4. *Enforcement*

A public prosecutor is authorized to bring a civil action against an operator in violation and seek civil penalties and other specified remedies. In addition, a child who suffers actual harm as a result of a violation, or a parent or guardian acting on behalf of that child, may bring a civil action against the operator to recover damages and other relief.

Any output provided by a companion chatbot that is in violation of the above provisions or any instance of an operator's failure to comply with a requirement constitutes a discrete violation.

#### 3. Stakeholder positions

The Los Angeles Unified School District writes in support:

With its comprehensive framework for mitigating a broad array of potential chatbot-linked harms to minors, AB 2023 institutes or expands on meaningful guardrails including age verification technology, parental controls and oversight, and restrictions on targeted advertising through conversational chats. The measure further adds assurances that interventions are working as intended by requiring the chatbot operators to submit to independent 3rd party audits and establishing clear and enforceable legal oversight by state and local entities.

As generative AI technology rapidly spreads into countless aspects of work, education, and daily life, unregulated chatbots pose significant risks to vulnerable users such as children, who are still developing the critical and analytical faculties adults take for granted. Because chatbots mimic human interaction children may perceive chatbot interactions as having human-like knowledge, intentions, and authority, which can lead them to trust and follow recommendations even when those systems produce inaccurate or harmful responses.

Writing in support, the California Catholic Conference argues:

AB 2023 takes meaningful and necessary steps to ensure that operators of companion chatbots are held accountable for protecting children. By requiring comprehensive risk assessments, independent audits, and reporting to the Attorney General, this bill promotes transparency, responsibility, and proactive safeguards against harmful content, including exposure to material related to self-harm or exploitation.

Catholic teaching calls us to uphold the dignity of the human person and to act in solidarity with the most vulnerable, particularly children who may be disproportionately impacted by emerging technologies. Young people are still developing the capacity for discernment and are especially susceptible to manipulation or harmful interactions in digital spaces. This bill reflects a commitment to the common good by ensuring that innovation does not outpace the moral responsibility to protect human life and the flourishing.

Writing in opposition, a coalition of groups, including the California Chamber of Commerce and the Computer and Communications Industry Association, identify a number of areas they would like to see amended, including defining the relevant harms, the liability structure, and the restrictions on selling, sharing, or using children's personal information. As to the auditing provisions, the coalition argues:

AB 2023's auditing provisions give rise to a host of concerns regarding unclear audit standards, a limited pool of qualified auditors, protections for sensitive business information, and broad AG disclosure authority, including to "qualified researchers". Without more objective statutory standards, audit findings may vary significantly across auditors, reducing their reliability and utility for both regulators and operators.

While third-party audits can play a role in compliance frameworks, they also impose significant costs and may not reliably advance safety objectives for AI systems. Audits are most effective when conducted against clear, objective standards, which are not generally suited to trying to evaluate a complex issue like the effects a platform has on minors. Here, auditors may be asked to evaluate highly subjective concepts such as emotional harm, autonomy, decisionmaking, and excessively sycophantic behavior. These are inherently difficult assessments that remain the subject of ongoing academic debate. Even academic researchers who have spent their entire careers studying the effects of technology use have not reached clear and convincing conclusions, given the impossibility of knowing the particular emotions and circumstances of an individual. In that context, these assessments are not readily susceptible to traditional audit methodologies and requiring third-party auditors to render definitive compliance judgments may produce inconsistent and unreliable results.

These challenges are compounded by the limited pool of qualified AI auditors and the significant costs associated with recurring audit obligations. For example, under European frameworks such as the Digital Services Act, companies have reported significant costs, limited auditor

availability, and continuous audit cycles. Given the nascent state of the AI auditing sector and the lack of widely accepted auditing standards, the Legislature should carefully consider whether the audit regime currently included in the bill is likely to provide the intended regulatory value.

#### 4. Amendment

One concern that has been raised is whether the obligations may interfere with certain opportunities older teens take advantage of in certain educational or workplace settings where companion chatbots have been incorporated. In order to avoid disrupting such opportunities, the author has agreed to an amendment that provides a narrow carveout from the definition of operator.

#### Amendment

(k) (1) "Operator" means a person who makes a companion chatbot available to a user in the state.

(2) "Operator" does not include either of the following:

(A) A postsecondary educational institution that makes a companion chatbot available exclusively for use in educational settings.

(B) An employer that makes a companion chatbot available exclusively to employees for use in workplace settings.

#### **SUPPORT**

Bright Light Strategies  
California Catholic Conference  
Los Angeles Unified School District  
Project Liberty LLC

#### **OPPOSITION**

American Innovators Network  
California Chamber of Commerce  
California Society of Certified Public Accountants (CALCPA)  
Civil Justice Association of California (CJAC)  
Computer & Communications Industry Association  
Insights Association  
Software Information Industry Association

#### **RELATED LEGISLATION**

SB 300 (Padilla, 2026) expands the protections in SB 243 by changing the threshold for when operators must carry out the obligations of the law to when they have actual

knowledge that a user is a minor. The bill also requires platform operators to prevent their chatbots from producing or facilitating the exchange of material of sexually explicit conduct or directly stating that the minor should engage in sexually explicit conduct, as provided. SB 300 is currently in the Assembly Privacy and Consumer Protection Committee.

SB 867 (Padilla, 2026) provides that no person shall manufacture, sell, or exchange, possess with intent to sell or exchange, or expose or offer for sale or exchange to any retailer any toy, as defined, that includes a companion chatbot. The bill sunsets on January 1, 2031. SB 867 is currently in the Assembly Privacy and Consumer Protection Committee.

SB 1119 (Padilla, 2026) is identical to this bill. SB 1119 is currently in the Assembly Judiciary Committee.

AB 1988 (Pellerin, 2026) requires operators that make companion chatbots available to have a policy and monitoring system for identifying and responding to “credible crisis expressions” in user conversations and details required responses to such expressions, including a crisis interruption pause after multiple such expressions. AB 1988 is set to be heard in this Committee the same day as this bill.

SB 243 (Padilla, Ch. 677, Stats. 2025) *See* Comment 2.

AB 1064 (Bauer-Kahan, 2025) would have prohibited making companion chatbots available to children unless the chatbots are not foreseeably capable of certain conduct, such as encouraging the child to harm others or themselves or engaging in sexually explicit interactions. Governor Newsom vetoed the measure, despite his acknowledgment that the “types of interactions that this bill seeks to address are abhorrent.” The Governor stated, in part: “While I strongly support the author’s goal of establishing necessary safeguards for the safe use of AI by minors, AB 1064 imposes such broad restrictions on the use of conversational AI tools that it may unintentionally lead to a total ban on the use of these products by minors. AI is already shaping the world, and it is imperative that adolescents learn how to safely interact with AI systems. This extends far beyond knowing how to use technology tools, such as conversational chatbots, and includes an understanding of what AI is, how it functions, and how to critically evaluate AI-generated content for algorithmic bias, misinformation, and other risks. We cannot prepare our youth for a future where AI is ubiquitous by preventing their use of these tools altogether.”

**PRIOR VOTES:**

Assembly Floor (Ayes 66, Noes 8)

Assembly Appropriations Committee (Ayes 12, Noes 3)

Assembly Privacy and Consumer Protection Committee (Ayes 13, Noes 2)

\*\*\*\*\*