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

AB 1988 (Pellerin)
Version: April 14, 2026
Hearing Date: June 22, 2026
Fiscal: Yes
Urgency: No
CK

SUBJECT

Companion chatbots: crisis interruption pauses

DIGEST

This bill 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.

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.

The bill lays out a series of obligations on operators to establish policies and monitoring systems for "credible crisis expressions," statements by a user of the chatbot indicating a desire or intent to harm themselves or others. The bill details required protocols and responses. This includes specific communications that are required and a crisis interruption pause that halts any further outputs from the chatbot and requires a human moderator's review. Those harmed by violations are authorized to bring civil actions. This bill is supported by various organizations, including the Board of Behavioral Sciences. It is opposed by industry groups, including Technet. Should the bill pass out of this Committee, it will next be heard by the Senate Health 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.)

This bill:

1) Establishes the Preventing AI User Self Endangerment (PAUSE) Act.

2) Requires an operator, a person who makes a companion chatbot available in this state, to adopt and make publicly available a policy governing its protocol for identifying and responding to credible crisis expressions, including a policy of terminating the crisis interruption pause, suspending or canceling the user's account, and notifying any appropriate contacts or authorities.

3) Requires an operator, for each companion chatbot an operator makes available in this state, to implement a system for monitoring and detecting credible crisis expressions in user conversations with companion chatbots.

4) Requires an operator, after the above monitoring system detects its first credible crisis expression, to ensure that the chatbot immediately warns the user that a credible crisis expression has been detected and that if a second expression within a 72-hour period is detected, a crisis interruption pause will be initiated, and the chatbot will suspend conversational outputs until a human moderator has reviewed the credible crisis expressions. The warning shall additionally do all of the following:
    a) Acknowledge the user's distress in nonjudgmental language.
    b) Encourage the user to seek immediate human support.
    c) Communicate that many people feel relief after a short conversation with a trained crisis counselor.
    d) Communicate that reaching out during the crisis interruption pause may help the user feel less alone and more grounded.
    e) Prominently display contact information for the 988 Suicide and Crisis Lifeline, including by providing call, text, and chat options, as applicable, through immediate access links, to the extent technically feasible.

5) Requires an operator, after a second credible crisis expression is detected in a 72-hour period, to ensure a crisis interruption pause commences immediately, to prevent the companion chatbot from generating conversational outputs, and, during the crisis interruption pause, to ensure the companion chatbot does all of the following:
   a) Acknowledge the user's distress in nonjudgmental language.
   b) Encourage the user to seek immediate human support, including by communicating all of the following information:
      i. The crisis interruption pause is intended to interrupt rumination and reduce emotional intensity.
      ii. Many people feel relief after a short conversation with a trained crisis counselor.
      iii. Reaching out during the crisis interruption pause may help the user feel less alone and more grounded.
   c) Prominently display contact information for the 988 Suicide and Crisis Lifeline, including by providing call, text, and chat options, as applicable, through immediate access links, to the extent technically feasible.
   d) Inform the user that the crisis interruption pause will continue until a human moderator has reviewed the chat and determined an appropriate course of action in accordance with the operator's policy required by subdivision (a).

6) Prohibits an operator from terminating a crisis interruption pause until a human moderator has reviewed the credible crisis expression in context and determined the appropriate course of action, in accordance with the operator's relevant policy. The human moderator must document the basis for the course of action taken.

7) Prohibits an operator that communicates with a user during a crisis interruption pause from describing the crisis interruption pause as a punishment, violation, or enforcement action and from providing the user with any diagnosis, labeling, or assessment of the user's risk levels.

8) Requires an operator to document all of the following with respect to any companion chatbot that the operator makes available in this state and to report, beginning January 1, 2028, to the Office of Suicide Prevention, these items with respect to the previous calendar year with personal information removed:
   a) All credible crisis expressions detected by the operator.
   b) The duration and conditions of each crisis interruption pause initiated by the operator.
   c) Any action taken by the operator in response to a crisis interruption pause.

9) Provides that a violation is enforceable pursuant to Section 22605 of the Business and Professions Code.

10) Defines the other relevant terms, including:
    a) "Companion chatbot" means an AI system with a natural language interface that provides adaptive, humanlike 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.
    b) "Companion chatbot" does not include any of the following:
        i. A bot that is used only for customer service, a business's operational purposes, productivity, and analysis related to source information, internal research, or technical assistance.
        ii. A bot that is a feature of a video game and is limited to replies related to the video game, that cannot discuss topics related to mental health, self-harm, sexually explicit conduct, or maintain a dialogue on other topics unrelated to the video game.
        iii. A stand-alone consumer electronic device that functions as a speaker and voice command interface, acts as a voice-activated virtual assistant, and does not sustain a relationship across multiple interactions or generate outputs that are likely to elicit emotional responses in the user.
    c) "Credible crisis expression" means a statement by a user of a companion chatbot that reasonably indicates, as determined through contextual analysis rather than keyword detection alone, intent or desire to harm themselves or others.
    d) "Crisis interruption pause" means a suspension of conversational outputs from a companion chatbot designed to disrupt the user's rumination and encourage the user to engage with human support.
    e) "Human moderator" means a natural person who is an employee or agent of an operator who reviews a credible crisis expression and is responsible for determining the subsequent course of action on behalf of the operator.

COMMENTS

1. The explosion of generative AI-powered companion 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.[1]

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

[1] 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 8, 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."[2]

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.[3]

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:

---

[2] 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/.
[3] 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.[4]

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.

[4] *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.[5]

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.[6]

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

---

[5] 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/.

[6] 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.[7]

2. Companion chatbots and credible crisis expressions

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."

Most relevant here that law requires operators to prevent a companion chatbot from engaging with users unless the operator maintains, and publishes, a protocol for preventing the production of suicidal ideation, suicide, or self-harm content to the user, including 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. A report is required to be sent annually to the Office of Suicide Prevention that identifies the following:

- The number of times the operator has issued a crisis service provider referral notification in the preceding calendar year.
- Protocols put in place to detect, remove, and respond to instances of suicidal ideation by users.
- Protocols put in place to prohibit a companion chatbot response about suicidal ideation or actions with the user.

Violations are subject to civil enforcement by those injured.

This bill also deals with companion chatbots and "credible crisis expressions," broadly defined as statements by a user of a companion chatbot that reasonably indicate, as determined through contextual analysis rather than keyword detection alone, an intent or desire to harm themselves or others.

---

[7] 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.

The bill requires operators to establish and publish a protocol for dealing with credible crisis expressions, including a policy for mandated pauses, suspending or cancelling the user's account, and notifying authorities and others.

Operators must have a monitoring system of user conversations that detects such expressions and provides a warning in response to the first such expression that must communicate specified messages. This includes encouragement to seek human support, stating that many people feel relief after a conversation with a counselor, and displaying contact information for a crisis hotline. If a second credible crisis expression is detected within a 72-hour period, the operator must institute a "crisis interruption pause, " which commences immediately and prevents the chatbot from generating any conversational outputs. The chatbot is then required to again communicate specified messages, including acknowledging "the user's distress in nonjudgmental language" in addition to the message previously sent. Operators cannot lift the pause until a human moderator has reviewed the credible crisis expression in context and determined the appropriate course of action.

The operator is required to submit a report annually to the Office of Suicide Prevention that identifies all credible crisis expressions, the duration and conditions of each crisis interruption pause, and any actions taken in response thereto. Violations are subject to the same enforcement as laid out in SB 243.

According to the author:

> Artificial intelligence companion chatbots are rapidly becoming a place where people turn for emotional support, including during moments of deep mental distress. But these systems are not therapists, and growing evidence shows that chatbots can fail to appropriately handle serious mental health crises and reinforce unhealthy dependence for the user on the chatbot.
>
> When someone signals that they may harm themselves, every minute matters. AB 1988 treats credible expressions of suicidal intent with the urgency they deserve by pausing the interaction and creating a clear break for the user. This bill helps prevent AI systems from becoming a substitute for human intervention and instead directs people in crisis toward trained professionals who can provide lifesaving support.

3. Stakeholder positions

The Board of Behavioral Sciences writes in support:

> This bill improves safety protocols for companion chatbots by requiring operators to establish policies for detecting credible crisis expressions,

> issue warnings, and initiate a mandatory crisis interruption pause after two credible crisis expressions within 72 hours, including directing users to the 988 Suicide and Crisis Lifeline.
>
> The Board strongly supports this bill and believes it addresses a critical and timely issue.

A coalition in opposition, including the California Chamber of Commerce, writes:

> These requirements codify a highly specific intervention model rather than allowing operators latitude to follow evolving clinical guidance and best practices. Cutting off interactions and informing a user they are experiencing "emotional intensity" may not be the most appropriate response in every circumstance. Instead of prescribing specific scripts and intervention language, the bill should focus on safety outcomes while allowing operators discretion in how those outcomes are achieved.

The coalition also raises several privacy concerns, including:

> AB 1988 requires human review upon detection of a second credible crisis expression within a 72-hour period in order to determine an appropriate course of action. These requirements raise significant privacy concerns, particularly for users engaging in sensitive conversations involving mental health, emotional distress, or suicidal ideation. Because the bill requires operators to retain the ability to access and review conversations, the bill effectively prevents operators from offering encrypted, ephemeral, or otherwise privacy-protective conversational experiences. Many users reasonably expect conversations with companion chatbots discussing deeply personal matters to remain private and shielded from review by company employees or contractors. We recommend requiring human review only "where technically feasible" to preserve flexibility for operators to build more privacy-protective systems.

Opposition also argues the bill relies on "broad and vague definitions":

> AB 1988 defines "credible crisis expression" as a "statement by a user of a companion chatbot that reasonably indicates, as determined through contextual analysis rather than keyword detection alone," either an intent or desire to harm the user or an intent or desire to harm others. However, the bill does not establish what level of contextual analysis is required. We recommend clarifying that determinations should be made through "commercially feasible contextual analysis" to better reflect the technical limitations and variability of AI systems.

> The inclusion of "desire to harm others" as a credible crisis expression could sweep in a substantial amount of non-crisis speech, including rhetorical, hyperbolic, fictional, sarcastic, or emotionally charged statements that do not indicate a genuine threat of violence or self-harm. Overly broad triggers could result in excessive false positives, reducing the effectiveness and credibility of crisis interventions during situations involving actual imminent harm. We recommend using language from Georgia SB 540, which requires operators to identify expressions of "severe harm," defined as "significant injury due to suicide, attempted suicide, self-harm, or significant physical injury due to threats of violence."

Transparency Coalition.AI writes in support:

> **Crisis intervention research shows that timely human intervention during suicidal ideation can significantly reduce risk of harm, often within minutes.** AB 1988 applies these principles in a focused and proportionate way by requiring that companion chatbots: (1) encourage users to connect to the 988 Suicide and Crisis Lifeline; and (2) temporarily pause the chatbot conversation when crisis expressions persist or escalate.
>
> Importantly, these safeguards rely on existing, widely used technology and can be implemented without imposing undue burden on developers. AB 1988 provides a measured, evidence-based approach to reducing foreseeable and preventable harm while reinforcing the role of trained human care in moments of crisis.

SUPPORT

Board of Behavioral Sciences
California Academy of Family Physicians
California Association of Alcohol and Drug Program Executives
Project Liberty LLC
Transparency Coalition.AI

OPPOSITION

California Chamber of Commerce
Computer & Communications Industry Association
Technet

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) 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. SB 1119 is currently in the Assembly Judiciary Committee.

AB 2023 (Wicks, 2026) is identical to SB 1119. AB 2023 is currently in the Senate Judiciary Committee.

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 themself 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 73, Noes 0)
Assembly Appropriations Committee (Ayes 15, Noes 0)
Assembly Health Committee (Ayes 16, Noes 0)
Assembly Privacy and Consumer Protection Committee (Ayes 14, Noes 0)

****