SB 300 (Padilla)
Version: January 5, 2026
Hearing Date: January 13, 2026
Fiscal: Yes
Urgency: No
CK

SUBJECT

Companion chatbots

DIGEST

This bill bolsters the law governing AI companion chatbots.

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. In response to these increasing concerns about their effects on users, including impacts on mental health and real-world relationships, SB 243 (Padilla, Ch. 677, Stats. 2025) was signed into law. It imposes a number of obligations on operators of "companion chatbot platforms" in order to safeguard users. This bill now 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.

This bill is author-sponsored and supported by the California Behavioral Health Association and others. Several industry groups are opposed, including TechNet.

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 artificial intelligence 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) 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) Provides that the relevant obligations on operators apply when the operator has actual knowledge that a user is a minor.

2) Requires operators to prevent their companion 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.

COMMENTS

1. <u>The explosion of generative AI-powered chatbots</u>

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 December 16, 2025.

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 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:

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

---

[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/.

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.[3]

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.

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.[4]

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

---

[3] *GPT-4o System Card* (August 8, 2024) OpenAI, https://openai.com/index/gpt-4o-system-card/.
[4] 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/.

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 further fortifies that law. It imposes these obligations on operators of these chatbot platforms when an operator has actual knowledge that a user is a minor. In addition, rather than simply require operators to institute reasonable measures to prevent the chatbots from producing sexually explicit material or directly stating a minor should engage in such conduct, it affirmatively requires 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.

This comes in the wake of additional reporting that internal documents at Meta indicate that their internal policy allowed for its chatbots to "engage a child in conversations that are romantic or sensual," generate false medical information, and help users argue that Black people are "dumber than white people."[5]

xAI, owned by Elon Musk, recently unveiled several sexually explicit chatbot companions on its platform.[6] For its part, OpenAI has announced plans to allow a "wider range of content, including erotica, on its popular chatbot ChatGPT."[7] This comes after a lawsuit against OpenAI was filed by the parents of Adam Raine, a teen who earlier last year committed suicide, accusing it of actively helping their son explore suicide methods:

> Adam's parents say that he had been using the artificial intelligence chatbot as a substitute for human companionship in his final weeks, discussing his issues with anxiety and trouble talking with his family, and that the chat logs show how the bot went from helping Adam with his homework to becoming his "suicide coach."[8]

---

[5] Jeff Horwitz, *Meta's AI rules have let bots hold 'sensual' chats with kids, offer false medical info* (August 14, 2025) Reuters, https://www.reuters.com/investigates/special-report/meta-ai-chatbot-guidelines/.
[6] Kate Conger, *Elon Musk Gambles on Sexy A.I. Companions* (October 6, 2025) The New York Times, https://www.nytimes.com/2025/10/06/technology/elon-musk-grok-sexy-chatbot.html.
[7] Lily Jamali & Liv McMahon, *ChatGPT will soon allow erotica for verified adults, says OpenAI boss* (October 15, 2025) BBC, https://www.bbc.com/news/articles/cpd2qv58yl5o.
[8] Angela Yang, et al., *The family of teenager who died by suicide alleges OpenAI's ChatGPT is to blame* (August 26, 2025) NBC News, https://www.nbcnews.com/tech/tech-news/family-teenager-died-suicide-alleges-openais-chatgpt-blame-rcna226147.

According to the author:

> Last year, I authored SB 243, which set necessary and vital guardrails around Artificial Intelligence companion chatbot technology. This bill was a critical first step towards making AI chatbots safer. SB 300 continues the valuable work done last year by strengthening the protections for minor users against AI exposing them to sexually explicit material. As AI technology rapidly evolves, we must ensure that we have the strongest safeguards in place to protect the most vulnerable among us, our children.

The California Initiative for Technology and Democracy writes in support:

> Many chatbots are designed to optimize user engagement, with the end product being an immersive "conversation" that becomes difficult to distinguish from a human-to-human interaction. This immersion makes it difficult to accurately and objectively judge the information provided by these tools, exacerbating an already dangerous misinformation problem that is inherent in the design.2 These products also have dangerous implications for the social wellbeing of our society. One recent study published by MIT and OpenAI found that heavy use of ChatGPT was correlated with increased loneliness, emotional dependence, and reduced socialization. These products have already had a particularly harmful impact on children, exacerbating an already daunting mental health crisis afflicting our country.
>
> SB 243 helped address this issue by requiring basic guardrails for companion chatbots to mitigate the mental health impacts of these products. SB 300 would strengthen these protections by requiring operators to actually prevent (rather than simply implement reasonable measures to prevent) a chatbot from exposing a minor user to sexually explicit material of any kind or encouraging the facilitation of sexually explicit content.

A coalition of industry groups, including TechNet, opposes the bill:

> SB 300 would replace key reasonableness-based standards adopted in SB 243 with a mandate to prevent the production of prohibited material and to broaden prohibited conduct to include "facilitating the exchange" of sexually explicit material. These are not clarifying or technical changes; they fundamentally alter the legal standard adopted by the legislature.
>
> At this stage, there is no demonstrated need for these changes and no clear gaps that the proposed language would fill. It is concerning to amend and expand statutory obligations before the Legislature, regulators, courts,

companies, or other stakeholders have had any opportunity to observe how the original law operates in practice.

SUPPORT

California Behavioral Health Association
California Initiative for Technology and Democracy
Transparency Coalition.ai

OPPOSITION

California Chamber of Commerce
Computer and Communications Industry Association
Technet

RELATED LEGISLATION

Pending Legislation: None known.

Prior Legislation:

SB 243 (Padilla, Ch. 677, Stats. 2025) *See* Executive Summary & 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."

****