ASSEMBLY THIRD READING
AB 1988 (Pellerin)
As Amended April 14, 2026
Majority vote

SUMMARY

Requires chatbot operators to ensure that multiple statements within a 72-hour period of a user′s intent to harm themself or others results in a suspension of the user′s account, pending human review.

Major Provisions

1) Defines, among other terms:

   a) ″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, intent or desire to harm themself or others.

   b) ″Crisis interruption pause″ as 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.

   c) ″Operator″ as a person that makes a companion chatbot available in this state.

   d) ″Human moderator″ as a human that 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.

2) Requires an operator to adopt and make publicly available a policy governing its protocol for identifying and responding to credible crisis expressions. Actions taken in accordance with the policy may include, but are not limited to, terminating the crisis interruption pause, suspending or cancelling the user′s account, and notifying any appropriate contacts or authorities.

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

4) If the monitoring system detects a credible crisis expression, requires the operator to do the following.

   a) For the first credible crisis expression, ensure that the chatbot immediately warns the user that a credible crisis expression has been detected and that if a second credible crisis 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 has reviewed the credible crisis expressions. Specifies that the warning must also do the following:

      i) Acknowledge the user′s distress in nonjudgmental language.

      ii) Encourage the user to seek immediate human support.

iii) Communicate that many people feel relief after a short conversation with a trained crisis counselor.

iv) Communicate that reaching out during the crisis interruption pause may help the user feel less alone and more grounded.

v) Prominently display contact information for the 988 Suicide and Crisis Lifeline, including by providing call, text, and chat options, as applicable. These options shall be made available to the user through immediate access links, to the extent technically feasible.

b) For the second credible crisis expression in a 72-hour period, ensure a crisis interruption pause commences immediately and prevent the companion chatbot from generating conversational outputs. During a crisis interruption pause, the operator must display a message with similar content to the warning that additionally informs the user that:

i) The purpose of the crisis interruption pause is to interrupt rumination and reduce emotional intensity.

ii) 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 described above.

5) 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 policy described above. Requires the human moderator to document the basis for the course of action taken.

6) Prohibits an operator that communicates with a user during a crisis interruption pause 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.

7) Beginning January 1, 2028, requires an operator to annually report to the Office of Suicide Prevention on crisis interruption pauses with respect to the previous calendar year. An operator shall ensure that the report does not contain any personal or identifying information of a user or other individual.

COMMENTS

Background. A growing body of evidence indicates that adaptive, sycophantic ″companion″ chatbots can create powerful feelings of attachment and trust among users – particularly those with vulnerabilities – by ensconcing them in a feedback loop that reinforces maladaptive beliefs. Mental health practitioners have encountered cases in which companion chatbots appear to have deteriorated the mental health of some individuals, leading to cases of delusions, self-harm, suicide, and harm to others. Top artificial intelligence companies are now facing a wave of lawsuits from aggrieved families who allege that chatbots were a substantial contributing factor in their loved ones taking their own lives or those of others.

This bill, the Protective AI Use Safety and Escalation (PAUSE) Act, would require operators of companion chatbots to implement a system for addressing ″credible crisis expressions″ –

statements that reasonably indicate an intent to harm one′s self or others. Under the bill, when a credible crisis expression is detected, the operator of the chatbot must ensure the chatbot displays a warning to the user that, among other things, encourages the user to seek human support and refers them to the 988 Suicide and Crisis Lifeline. If a second credible crisis expression is detected within 72 hours, the operator must initiate a crisis interruption pause during which no further conversational outputs are allowed until a human moderator reviews the credible crisis expression in context and determines and documents the appropriate course of action in accordance with the operator′s policy. The bill also requires operators to submit an annual report to the Office of Suicide Prevention with specified information related to the implementation of the bill.

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 or others, 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.

Arguments in Support
Didi Hirsch Mental Health Services, the bill′s sponsor, writes:

> Californians, and individuals worldwide, are increasingly relying on companion chatbots for support and advice during moments of acute psychological distress. *However, these tools are not equipped to deliver appropriate care to people in crisis, creating a clear and growing public safety gap between how they are used and what they are capable of providing.*
>
> A pattern of recent incidents underscores the potentially devastating effects of the current lack of guiding legislation around AI companion chatbots. In 2025, the family of a 16-year-old boy filed a wrongful death lawsuit alleging that a chatbot validated his suicidal ideation, assisted in drafting a suicide note, and failed to direct him to human support before his death. *Additional lawsuits allege that chatbots have, in some cases, provided detailed guidance on methods of self-harm or suicide, underscoring the risk of systems responding in ways that may actively exacerbate harm rather than interrupt it.*
>
> Emerging research reinforces that these incidents are part of a broader pattern of harm. A 2025 Stanford study found that chatbots can generate inappropriate and harmful responses when users present with serious mental health conditions, including suicidal ideation, and research from Brown University found that these systems can fail to adhere to established standards of clinical care, even when prompted to use evidence-based psychotherapy techniques.
>
> *Crisis intervention research shows that timely human intervention during suicidal ideation can significantly reduce risk of harm, often within minutes.* [. . .] (Emphasis in original.)

Arguments in Opposition
In opposition to the bill, California Broadband & Video Association argues that the bill's definition of "companion chatbot" – drawn from existing law – is too broad and instead recommends adopting a definition of "companion chatbot" from New York law:

> To ensure AB 1988 focuses on the specific category of applications that raise legitimate concerns, we respectfully recommend aligning the bill's definition with a similar law enacted last year in New York. That framework more clearly distinguishes AI systems designed to simulate sustained emotional relationships from general-purpose AI tools.
>
> In particular, the New York approach emphasizes sustained personal dialogue and unprompted emotional engagement, which helps differentiate higher-risk AI companions from standard productivity or information tools.
>
> We respectfully request that AB 1988 define "Artificial Intelligence Companion" as:
>
> "Artificial Intelligence Companion" means a software application that uses generative artificial intelligence and is designed, marketed, or optimized to simulate a sustained human or human-like social or emotional relationship with a user by: (A) retaining information from prior interactions or user sessions to personalize ongoing engagement; (B) asking unprompted or unsolicited emotion-based questions that go beyond responding to a direct user prompt; and (C) sustaining ongoing dialogue concerning matters personal to the user.

FISCAL COMMENTS

According to the Appropriations Committee:

> Negligible costs to [California Department of Public Health].
>
> Cost pressures (Trial Court Trust Fund, General Fund) of an unknown but potentially significant amount to the courts to adjudicate any additional filings. Actual costs will depend on the number of cases filed and the amount of court time needed to resolve each case. It generally costs approximately $1,000 to operate a courtroom for one hour. Although courts are not funded on the basis of workload, increased pressure on the Trial Court Trust Fund may create a demand for increased funding for courts from the General Fund. The state budget provides annual General Fund backfills to the Trial Court Trust Fund to offset revenue reductions, totaling approximately $117.3 million in 2025-26.

VOTES

ASM PRIVACY AND CONSUMER PROTECTION: 14-0-1
YES: Bauer-Kahan, Macedo, Aguiar-Curry, Bryan, DeMaio, Hoover, Irwin, Lowenthal, McKinnor, Ortega, Petrie-Norris, Ward, Wicks, Wilson
ABS, ABST OR NV: Patterson

ASM HEALTH: 16-0-0
YES: Bonta, Chen, Addis, Aguiar-Curry, Ahrens, Caloza, Carrillo, Mark González, Johnson, Patel, Patterson, Rogers, Sanchez, Schiavo, Sharp-Collins, Stefani

ASM APPROPRIATIONS: 15-0-0
YES: Wicks, Hoover, Aguiar-Curry, Calderon, Caloza, Dixon, Fong, Mark González, Krell, Pacheco, Pellerin, Sharp-Collins, Solache, Ta, Tangipa

UPDATED

VERSION: April 14, 2026

CONSULTANT: Josh Tosney / P. & C.P. / (916) 319-2200 FN: 0002951