ASSEMBLY THIRD READING
AB 2023 (Wicks and Bauer-Kahan)
As Amended April 27, 2026
Majority vote

## SUMMARY

Establishes a comprehensive regulatory framework to ensure that companion chatbots made available to children in this state are safe by design.

**Major Provisions**

1) Requires chatbot operators to implement age verification requirements pursuant to AB 1043 (Wicks), Chapter 675, Statutes of 2025 or apply specified protections under the bill to all users.

2) Requires an operator to do all of the following by July 1, 2027:

   a) Perform and document a comprehensive risk assessment to identify child safety risks posed by the design, configuration, and operation of the companion chatbot that assesses: the likelihood of a covered harm occurring to users; differential risks across age groups and developmental stages; known vulnerabilities of children; empirical data from actual use; relevant academic research and regulatory guidance.

   b) Take and document measures to reasonably mitigate identified risks.

   c) Publish a child safety policy on its internet website, and update the policy as necessary to ensure accuracy.

   d) Implement:

      i) A documented crisis response protocol to mitigate any risk that the companion chatbot will generate a statement that promotes suicidal ideation, suicide, or self-harm content, including, but not limited to:

         (1) Timely in-service support and clear referral to appropriate external crisis resources if the operator determines a child has expressed suicidal ideation or self-harm.

         (2) If a child′s account is connected to a parent′s account, default notifications to the parent within 24 hours if the child′s account shows a substantial risk that the child may suffer a covered harm.

         (3) Clear and age-appropriate disclosures to child users whose accounts are linked to a parent′s account that inform them that a parent may be notified if the companion chatbot detects content or behavior that indicates potential risks to the child′s safety or well-being.

      ii) Safeguards for child users that include usage reminders and disclosures, age-appropriate risks prompts, and other protective design features reasonably related to documented child safety risks.

iii) Default settings that can be changed only by a parent that include the following:

(1) Setting the chatbot to ephemeral mode, unless the parent provides affirmative consent for persistent conversational memory.

(2) A prohibition on push notifications during specified hours.

(3) Limiting the amount of time a child can spend on a single conversation with the chatbot to one hour, and the total time to two hours per day.

iv) A mechanism for providing notice to a child user that the child is interacting with, or receiving content generated by, an AI system, as specified.

v) Measures that prevent the companion chatbot from encouraging the child to engage in specified harmful behaviors; attempting to diagnose or treat a user′s physical, mental, or behavioral health, except as specified; engaging or depicting an individual in obscene matter or sexual abuse material with a user; discouraging a child from sharing health or safety concerns with a qualified professional or adult; discouraging the child from taking breaks or suggesting the child needs to return frequently; claiming that the chatbot is sentient, conscious, or human; soliciting gifts or other expenditures; facilitating product advertisements during chats; producing responses that are excessively sycophantic.

vi) Parental controls that are accessible, easy-to-use that can be connected to a child′s account, including those that allow for control of whether and to what extent the companion chatbot uses persistent conversational memory, setting preferences for the chatbot′s interaction with the child, setting time limits for the child′s use of the companion chatbot; and disabling access for children under 16 years of age. Operators must promote parental controls through reasonable communication methods and provide prompt notice to a parent if the child modifies or disables a parental control.

vii) Interface designs that ensure features and controls are accessible and clear, so that children and parents can reasonably locate, understand, and use those protections. Operators must annually test the interface, as specified.

viii) A public incident reporting mechanism that enables a third party to report directly to the operators an incident regarding a child safety risk and to access high-level summaries of other reports made through that reporting mechanism.

3) Prohibits operators from:

a) Targeting advertising to a child, as specified.

b) Selling, sharing, or using a child′s personal information for any purpose not expressly authorized by the bill.

c) Using dark patterns to prevent users from being able to use features and controls required under the bill.

4) Requires operators to submit to annual independent audits, beginning 180 days after the AG adopts regulations. Audit reports must be submitted to the AG, but must be kept confidential. The AG may disclose audit reports to government agencies and public prosecutors for enforcement purposes and researchers, subject to confidentiality agreements.

5) Requires the AG, by January 1, 2028, to:

   a) Adopt regulations governing audits.

   b) Establish a public incident reporting mechanism.

   c) Establish a process for qualified researchers to access anonymized and aggregated audit data for academic study of child safety in companion chatbots.

6) Beginning January 1, 2028, requires the AG to issue an annual report that includes specified information about audits and recommendations for operators, parents, and policymakers.

7) Authorizes public prosecutors to bring an action against violators for a civil penalty of up to $5,000 per negligent violation, per child, and up to $15,000 per intentional violation, per child, as well as for punitive damages, injunctive or declaratory relief, reasonable attorney′s fees, and any other relief the court deems proper.

8) Allows minors who suffer actual harm, or parents on their behalf, to bring a civil action against violators for actual damages, punitive damages, injunctive or declaratory relief, reasonable attorney′s fees, and any other relief the court deems proper.

9) Provides that the duties, remedies, and obligations imposed by the bill are cumulative to those elsewhere in the law.

10) Contains a severability clause.

## COMMENTS

*Background.* Last year′s SB 243 (Padilla) requires chatbot platforms to establish protocols to detect, remove, and respond to instances of suicidal ideation, suicide, or self-harm expressed by users. For users that are minors, SB 243 further requires operators to disclose to users that they are interacting with AI, provide periodic reminders to take a break and that the chatbot is artificially generated, and prevent chatbots from producing sexually explicit material.

A complementary bill, AB 1064 (Bauer-Kahan, 2025), would have prohibited making available to minors a companion chatbot that is foreseeably capable of specified harmful behaviors, including encouraging the child to engage in self-harm, suicidal ideation, or violence, or engaging in sexually explicit interactions with the child. Claiming that the bill could ″unintentionally lead to a total ban on the use of these products by minors,″ Governor Newsom vetoed the bill, stating:

> The types of interactions that this bill seeks to address are abhorrent, and I am fully committed to finding the right approach to protect children from these harms in a manner that does not effectively ban the use of the technology altogether. I will work with my partners in the Legislature to build on the framework established by SB 243 (Padilla) to develop a bill

> next year that ensures young people can use AI in a manner that is safe, age-appropriate, and in the best interests of children and their future.

Earlier this year, AB 1064′s sponsor, Common Sense Media, teamed up with OpenAI to introduce a ballot measure that would have enacted a comprehensive framework for regulating chatbots. The measure includes provisions relating to age assurance, risk assessments, content restrictions, privacy, audits, and enforcement. However, a coalition of child safety advocates, civil society groups, and technology policy organizations criticized the measure and the process by which it was proposed.[1] In their letter in support of this bill, Children Now states that the initiative ″contained numerous loopholes, partial protections, and serious limitations on the ability to enforce the law.″ Stakeholders who support and oppose this bill appear to agree that it is far preferable to address this issue through the standard legislative process. Common Sense Media and OpenAI have paused the effort, although the ballot committee remains open.

This bill and a parallel measure, SB 1119 (Padilla) of the current legislative session, seeks to build on the framework proposed in the initiative. The measures include provisions relating to age assurance, risk assessments, content restrictions, privacy, audits, and enforcement.

**According to the Author**
AB 2023 would establish a comprehensive framework to address the risk of chatbot interactions by children. Some of these guardrails would include: protocols to address suicidal ideation, sycophancy, and isolation; default settings for children; parental controls; noticing requirements; crisis response protocols; prohibitions on advertising and the selling, sharing; prohibition on the usage of children′s private information; robust oversight and enforcement framework including through a public incident reporting mechanism; third-party audits; the development of auditing standards by the attorney general; and including a private right of action.

**Arguments in Support**
The California Initiative for Technology & Democracy, which supports the bill, writes about tragic incidents involving chatbots, stating:

> These incidents are not isolated, and the widespread use of chatbots suggests the dangers could grow. Last year, it was reported that over 50% of students have used chatbots to help with homework, and 20% have engaged in a romantic relationship with AI. The underlying danger of these chatbots lies in how they interact with users. This process often rewards responses that affirm a person′s beliefs rather than challenge them, regardless of what would actually benefit the user. Consider a teenager who expresses to a chatbot that they feel worthless and want to disappear, rather than redirecting to crisis resources, a sycophantic model is incentivized to validate and deepen that emotional state because agreement generates positive feedback. This ″sycophancy″ is not merely a dangerous side effect; it is a curated design choice.
>
> Chatbots also raise serious privacy concerns, particularly for children. As users form emotional relationships with these tools, they are likely to disclose intimate details about their lives. This information can then be used for model training, enabling responses that feel especially personalized. More alarming is that many developers have already disclosed plans

[1] Open Letter, "Oppose OpenAI Writing Its Own Regulations in California," (March 18, 2026), https://whowritestherules.org/.

to use this data for advertising, taking users′ most intimate social, health, or sexual conversations and monetizing them for ad revenue. This is the ultimate commodification of intimacy, and it creates a powerful mechanism for manipulation.

AB 2023/SB 1119 would require an annual risk assessment along with the establishment of measures to prevent suicidal ideation, sycophancy, and isolation, including a crisis response protocol. It would provide added guardrails in the form of default settings for children, parental controls, notice requirements, and time limits. It would prohibit advertising and the sale, sharing, and use of children′s private information. And it would establish a robust oversight and enforcement framework, including a public incident reporting mechanism, third-party audits, auditing standards developed by the Attorney General, and a private right of action.

By instituting these safeguards, AB 2023/SB 1119 addresses many of the most pressing documented dangers of chatbots, including sycophancy, sexual entanglement, and self-harm. Importantly, it mandates strict default privacy settings, reducing the burden on parents who cannot realistically keep up with every new parental control. The bill also establishes meaningful accountability for the companies that produce these products, requiring independent third-party companies have profited while their users suffer, without consequence.

**Arguments in Opposition**
TechNet writes:

We appreciate the author′s most recent amendments, which, among other things, helped clarify certain definitions of harm to align with violations of existing law and clinical guidelines, clarified the prohibition on targeted advertising, created flexibility regarding the use of age verification, and reduced risks of exposing sensitive information through public reporting.

While these measures improve the bill, we maintain concerns about the bill′s interactions and conflicts with last year′s SB 243 (Padilla), overly prescriptive requirements, and overlapping enforcement requirements. We welcome the opportunity to continue working with the sponsors to address these concerns.

## FISCAL COMMENTS

According to the Appropriations Committee:

1) Costs (General Fund) of an unknown but potentially significant amount, likely in the low hundreds of thousands of dollars at a minimum, to the Department of Justice (DOJ) for one-time rulemaking to adopt regulations on or before January 1, 2028, governing third-party audits, including professional and ethical standards for auditors, auditor eligibility requirements, audit procedures, and audit report content. Additional unknown, but potentially significant costs, to DOJ for enforcement actions against operators for negligent or intentional violations, including investigatory workload and litigation costs. Actual costs will depend on whether the Attorney General pursues enforcement actions, and, if so, the level of additional staffing DOJ needs to handle the related workload. If DOJ hires staff to handle enforcement actions authorized by this bill, the department would incur significant costs, likely in the low hundreds of thousands of dollars annually at a minimum. If DOJ does not pursue enforcement as authorized by this bill, the department would likely not incur any

costs. Potential offsetting revenues (General Fund) from civil penalties collected pursuant to enforcement actions. The DOJ was unable to provide a cost estimate at the time this analysis was written.

2) Cost pressures (Trial Court Trust Fund, General Fund) of an unknown but potentially significant amount to the courts to adjudicate civil enforcement actions and cases filed under the private right of action created by this bill. 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.

## VOTES

**ASM PRIVACY AND CONSUMER PROTECTION: 13-2-0**
**YES:** Bauer-Kahan, Bryan, Hoover, Irwin, Lowenthal, McKinnor, Ortega, Patterson, Pellerin, Petrie-Norris, Ward, Wicks, Wilson
**NO:** Macedo, DeMaio

**ASM APPROPRIATIONS: 12-3-0**
**YES:** Wicks, Hoover, Aguiar-Curry, Calderon, Caloza, Fong, Mark González, Krell, Pacheco, Pellerin, Sharp-Collins, Solache
**NO:** Dixon, Ta, Tangipa

## UPDATED

VERSION: April 27, 2026

CONSULTANT: Josh Tosney / P. & C.P. / (916) 319-2200 FN: 0002963