VIEWPOINT

Mitigating Suicide Risk for Minors Involving AI Chatbots—A First in the Nation Law

I. Glenn Cohen, JD; Julian De Freitas, PhD

On October 13, 2025, California enacted a series of artificial intelligence (AI) regulation bills, including Senate Bill 243 (SB 243), which is the first law in the nation to regulate AI "companion chatbots."[1] Many users, including minors, have used such chatbots to help with their mental health, but it has been alleged in ongoing litigation that some systems may have contributed to cases of teen suicide.[2] Although the new law formally only governs in California—given that many of the AI operators are based there, the size of the state's user base, and prior experience with laws like the California Privacy Act—we believe it is likely to set the standard for AI chatbot regulation across the country and potentially the world. In this Viewpoint, we describe the California law, its positive contributions to making these companion chatbots safer for minor and adult users, but also the law's limits. We recommend further steps California or other states can adopt to improve protections for mental health and chatbot safety, especially for minors.

SB 243 governs companion chatbots, which it defines as any "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."[1] The statute explicitly excludes, among other things, chatbots used for customer service, research, some virtual assistants, and video games that "cannot discuss topics related to mental health, self-harm, sexually explicit conduct, or maintain a dialogue"[1] unrelated to the game.

SB 243 main requirements are, first, that if a "reasonable person" interacting with the chatbot would be "misled to believe that the person is interacting with a human" then the company must issue a "clear and conspicuous notification indicating that the companion chatbot is artificially generated and not human."[1] Second, the chatbot operator must publish on its website details about its "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."[1] Third, for "a user that the operator knows is a minor," the operator must disclose that the minor is interacting with an AI chatbot, provide a clear and conspicuous notice reminding the user that the chatbot is AI and not human after every 3 hours of interaction and put in "reasonable measures" to prevent the chatbot from producing sexually explicit material or telling the user to engage in sexually explicit conduct.[1] Fourth, starting in 2027 it requires operators to produce an annual report on the protocol, the number of notifications regarding suicidal ideation, and other data. Last, it permits civil lawsuits for the violation of the act, including for damages of at least $1000 per violation or greater penalties if the actual damages are higher.[1]

SB 243 would not seem to reach most uses of chatbots by pediatricians and health care professionals in the course of delivering treatment because its definition excludes a chatbot "that is used only for customer service" or "a business' operational purposes."[1] California should be applauded for (and other states should emulate) SB 243's requirements to inform users that a companion chatbot is AI, to issue periodic reminders of that fact, to recommend that minors take breaks after a set period of use, and to develop protocols for responding to signs of suicidal ideation and sharing of information about suicide hotlines and related resources. At the same time, we believe each of these elements could be improved in future legislation. It is not clear why the right to know one is dealing with an AI should be limited to companion chatbots instead of *all* chatbots. The law's notification requirement only attaches when the operator "knows" the user is a minor, but the law does not provide information about how to assess what knowledge the operator must have. Operators can design systems precisely not to be able to detect that information, such as ignoring whether the user discloses their age during the chat. A recent audit of 15 companion chatbots found that only 5% even specify a minimum age for use of their services.[2]

Operators should be required to assume a user is a minor unless they can demonstrate otherwise. The form of the required notification in SB 243 is unspecified and, without knowing more, we do not know whether operators will design the most effective forms of notice that might also discourage use time. The trigger for notification of "at least every three hours for continuing companion chatbot interactions,"[1] leaves ambiguous what it means for an interaction to be "continuing" and might permit more than 18 hours of interaction per day if punctuated by short breaks. A more defensible standard would track cumulative daily use, ideally paired with a requirement to monitor the emotional tone of interactions, because risk stems not only from duration but from content.

Requiring a protocol to be in place for preventing and detecting suicidal ideation is a meaningful start, but the statute stops short of defining what such a protocol must entail—or even requiring that it be effective or reviewed by suicide prevention experts. SB 243 has favored a procedural step over substantive requirements, and there may be a lot of heterogeneity in terms of what operators adopt. This procedural approach contrasts sharply with the more protective model envisioned in California's AB 1064, vetoed by its governor, which would have barred companies from offering chatbots that are foreseeably capable of encouraging self-harm, suicide, violence, substance use, or disordered eating or of providing unsupervised mental health therapy or erotic interactions with minors.[3] The governor

stated he vetoed AB 1064 because, even though he supported the "goal of establishing necessary safeguards for the safe use of AI by minors," he thought the bill imposed "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."[4] Still, AB 1064's focus on a standard for preventing foreseeable harms, not just documenting them, is something future legislation should return to.

SB 243's requirement of periodic reminders that users are speaking with an AI is unlikely to counteract the emotional realism of systems deliberately designed to appear human. Children can be told "this is not a human" and yet when their perceived experience tells them the exact opposite, it is hard for them to avoid ascribing the chatbot intentions, emotions, motivations and other characteristics that can engender trust and emotional attachment, which in turn can make them more vulnerable to abuse or changes that perturb this attachment.[5] Furthermore, even adults who say goodbye to a disclosed chatbot before intending to log off are more likely to stay engaged on the app if the chatbot uses emotional manipulation tactics such as fear of missing out or emotional neglect.

Rather than just a notice requirement, we would urge California (and other states) to build out legislative requirements to test for risks of emotional dependence and require operators to justify design aspects, including humanlike features, which lead to emotional manipulation or dependence for any chatbots that could interact with minors. A related area for further legislation is the line between virtual "friend" and "therapist." The vetoed AB 1064 would have restricted the chatbot offering "mental health therapy to the child," whereas the current law has eschewed policing that line. When it comes to human beings, states are serious about policing the unauthorized practice of medicine and require strict licensure rules to provide therapy.[6] It is unclear why an AI chatbot, whose responses are harder to predict and whose training data and architecture are opaque, should be exempt from comparable oversight—including mandatory reporting of suicidal ideation to a minor's parents—merely because it is virtual.

California's law is a genuine milestone. But if the aim is truly to reduce suicide risk and protect mental health, this should be seen as a foundation, not a finish line. The next challenge is ensuring that disclosure, safety protocols, and emotional safeguards are designed with evidence—not optics or box checking—in mind. Industry lobbying will no doubt dilute stricter standards, but policymakers should not mistake minimal compliance for meaningful protection—especially when the systems in question may be a child's most responsive voice in moments of distress.

ARTICLE INFORMATION

Author Affiliations: Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics, Harvard Law School, Harvard University, Cambridge, Massachusetts (Cohen); Marketing Unit, Harvard Business School, Harvard University, Boston, Massachusetts (De Freitas).

Corresponding Author: I. Glenn Cohen, JD, Harvard Law School, 1525 Massachusetts Ave, Cambridge, MA 02138 (igcohen@law.harvard.edu).

Published Online: December 22, 2025. doi:10.1001/jama.2025.23744

Conflict of Interest Disclosures: Mr Cohen reported receiving grants from the Novo Nordisk Foundation; being a member of the Bayer Bioethics Council; being a bioethics adviser to Bexorg, World Class Health, and Manhattan Neuroscience; chairing an ethics advisory board for Illumina; receiving speaking fees from Philips, the *Washington Post*, and the Doctors Company; attending a leadership forum organized by Galen/Atlantica; and being retained by [redacted] as an expert in health privacy, gender-affirming [redacted] and reproductive technology lawsuits. Dr De Freitas reported being a scientific adviser to Spring Health and Flourish Science.

REFERENCES

1. State of California Legislative Information website. SB-243 Companion chatbots. Accessed November 21, 2025. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB243

2. Brewster RCL, Zahedivash A, Tse G, Bourgeois F, Hadland SE. Characteristics and safety of consumer chatbots for emergent adolescent health concerns. *JAMA Netw Open*. 2025;8(10):e2539022-e2539022. doi:10.1001/jamanetworkopen.2025.39022

3. LegiScan. California Assembly Bill 1064. Accessed November 21, 2025. https://legiscan.com/CA/text/AB1064/id/3267991

4. Nguyen T. California governor vetoes bill to restrict kids' access to AI chatbots. Published October 15, 2025. Accessed November 21, 2025. https://apnews.com/article/california-chatbots-children-safety-ai-newsom-33be4d57d0e2d14553e02a94d9529976

5. De Freitas J, Cohen IG. Disclosure, humanizing, and contextual vulnerability of generative AI chatbots. *NEJM AI*. Published online January 21, 2024. doi:10.1056/AIpc2400464

6. De Freitas J, Cohen IG. The health risks of generative AI-based wellness apps. *Nat Med*. 2024;30(5):1269-1275. doi:10.1038/s41591-024-02943-6