

**To:** Kenichi Maruyama, PhD, MPA, GPC

**From:** Alizea Alvarez-Sookram

**Date:** 07/12/25

**Re:** Governing Emerging Technologies: Policy Options for AI Tools in Youth Mental Health

## Executive Summary

Artificial intelligence (AI) tools are showing up everywhere in youth mental health, from chatbots that simulate therapeutic conversations to early-warning systems that flag signs of distress. In a moment when millions of adolescents experience depression or suicidal ideation each year and school counselors are stretched impossibly thin, it's easy to see why these tools feel both promising and inevitable (Reinert et al., 2025; Inseparable, 2025).

As I reviewed the evidence, I kept coming back to a simple tension: AI is moving faster than our systems are prepared for. Many tools show potential, yet most haven't been rigorously validated, and several high-profile cases reveal how quickly things can go wrong, from chatbots encouraging self-harm to incidents of "AI psychosis" that raise red flags about safety and accountability (Harwell, 2024; Cha, 2025; Rauschenberg et al., 2021). Policymakers are responding, but their approaches are scattered and inconsistent.

This memo compares three policy paths:

- **Option A:** California's standards-based model for AI companions
- **Option B:** State bans and restrictions on AI therapy
- **Option C:** Federal public health guidance through the Surgeon General

After weighing the tradeoffs, I recommend **Option A** as the primary approach, supported by targeted uses of Options B and C. A standards-based framework provides the clearest path to safer tools without pushing youth toward unregulated alternatives, while federal guidance helps set national expectations. Bans remain necessary, but only in narrow, high-risk contexts.

## Problem Statement

Youth mental health needs are high, urgent, and growing. Nearly 2.8 million adolescents experienced a major depressive episode in 2024, and close to 3 million reported suicidal thoughts. These numbers feel less like statistics and more like a reflection of what's happening in schools and families across the country (Reinert et al., 2025).

Schools often have one counselor for 700+ students (Inseparable, 2025). In that gap, AI tools have flourished: automated risk-flagging systems, AI-assisted screening, and chatbots offering cognitive-behavioral support (Sharma et al., 2025; Li et al., 2025). These tools are not fringe, they're already present in classrooms, apps, and after-hours support.

Yet the risks are just as real: inappropriate responses, data privacy concerns, and emerging clinical questions about disorientation or "AI psychosis" (Cha, 2025). The challenge is not

whether AI will show up in youth mental health, really it already has, but how we govern it so that it actually supports young people rather than amplifying their vulnerabilities.

### **Policy Option 1: Safety Standards for AI Companions**

California's Assembly Bill 1064 offers a thoughtful blueprint for regulating AI companions marketed to minors. The bill requires key safeguards: crisis-response features that escalate to human hotlines, clear disclosures (this is a chatbot, not a human), and ongoing reporting about user impacts (California Legislative Information, 2025).

What I appreciate about this model is its practicality. Instead of pretending AI will disappear, it asks: *If young people are going to use these tools anyway, how do we make them safer?*

The government mechanism here is standards-setting: clear rules of the road that companies must meet before their products can be marketed to youth. It prioritizes transparency, accountability, and proactive risk mitigation.

### **Policy Option 2: State Bans and Restrictions on AI Therapy**

Other states have gone in the opposite direction. Illinois and Nevada have enacted bans on AI-driven therapy, arguing that unvalidated tools shouldn't substitute for clinicians when minors are involved. Utah takes a more middle-ground approach, allowing limited use with strict disclosures and privacy requirements (Schoenherr, 2025).

This model prioritizes a precautionary stance: if the tool might harm young people, err on the side of restriction. The instruments here are command-and-control regulation and mandatory disclosure.

The upside is clear: high-risk tools are kept out of therapeutic settings. The downside is equally important: bans often push youth toward unregulated online tools, far outside the reach of clinicians, schools, or policymakers.

### **Policy Option 3: Federal Public Health Guidance**

At the federal level, the U.S. Surgeon General has issued advisories highlighting the risks associated with digital platforms, including those using AI (Office of the U.S. Surgeon General, 2023). These advisories don't carry legal force, but they shape national conversations.

The power here is normative: federal guidance can set expectations for transparency, evaluation, and safety-by-design, which can then be woven into grant requirements, procurement decisions, and public-health campaigns. It creates a kind of gentle consistency across states without imposing one-size-fits-all rules.

### **Policy Recommendation**

After comparing the options, **Option A**, safety standards for AI companions, is the most balanced and forward-thinking path. It acknowledges that these tools are already widely used by

youth, while building in the protections that have been missing. It's grounded, feasible, and supported by evidence that regulation and disclosure improve safety without shutting down innovation (Sharma et al., 2025).

**Option C** is the ideal complement. Federal advisories help unify the message: AI tools must be transparent, safe-by-design, and continuously evaluated. Even without enforcement power, federal guidance raises the bar and signals what *responsible* looks like.

**Option B**, bans and restrictions, should remain a last resort. They are appropriate when the risks are unmistakably high. For example, tools that claim to offer therapy without evidence, or platforms that repeatedly produce unsafe responses. But bans should be narrowly tailored and ideally paired with exceptions for research and pilot programs so that we continue learning rather than freezing innovation prematurely.

**In summary, the ideal layered approach is:**

1. **Option A** - the foundation: safety standards that make existing tools safer.
2. **Option C** - national direction-setting and agenda shaping.
3. **Option B** - targeted restrictions for the highest-risk contexts only.

This combination provides a realistic, equity-minded path for protecting young people while supporting responsible, evidence-based innovation in AI-enabled mental health care.

## References

California Legislative Information. (2025). *AB-1064 Leading Ethical AI Development for Kids*

*Act*. [Bill Text - AB-1064 Leading Ethical AI Development \(LEAD\) for Kids Act](#).

Kraft, M. E., & Furlong, S. R. (2021). *Public policy: Politics, analysis, and alternatives* (7th ed.). CQ Press.

Office of the U.S. Surgeon General. (2023). *Advisory on social media and youth mental health*.

U.S. Department of Health & Human Services. [Social Media and Youth Mental Health | HHS.gov](#)

Schoenherr, N. (2025, September 13). *California bill aims to regulate companion AI for kids, amid national debate on AI therapy bans*. *The Hill*. [California AI Child Protection Bill Clears Statehouse](#)

Rauschenberg, C., Schick, A., Hirjak, D., Seidler, A., Paetzold, I., Apfelbacher, C.,

Riedel-Heller, S. G., & Reininghaus, U. (2021). Artificial intelligence in mental health care: A systematic review of diagnosis, monitoring, and intervention applications.

*Psychological Medicine*, 53(6), 1–20. [Artificial intelligence in mental health care: a systematic review of diagnosis, monitoring, and intervention applications | Psychological Medicine | Cambridge Core](#)