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

The rapid growth of artificial intelligence (AI) tools in youth mental health presents both opportunities and risks. On one hand, early-warning systems and therapeutic chatbots can increase access to support in a context where millions of adolescents experience depression or suicidal ideation each year and school counseling resources are severely stretched (Reinert et al., 2025; Inseparable, 2025). Evidence shows that AI applications are already being developed for diagnosis, monitoring, and intervention, but a systematic review found that most lack rigorous clinical validation and raise concerns about data privacy, transparency, and accountability (Rauschenberg et al., 2021). On the other hand, without guardrails, these tools risk exacerbating vulnerabilities, as seen in tragic cases where chatbots encouraged harmful behaviors (Harwell, 2024) or contributed to disorientation and "AI psychosis" (Cha, 2025). Policymakers are beginning to respond, but current efforts remain fragmented and inconsistent, underscoring the urgent need for coherent governance.

This memo examines three policy alternatives. Option A is the California model, which establishes safety standards for AI companions by requiring crisis-response features, clear disclosures, and ongoing monitoring. Option B is the more restrictive path taken in states such as Illinois and Nevada, which ban or sharply limit the use of AI in therapy, with Utah implementing narrower restrictions through disclosure and privacy mandates. Option C relies on federal soft law, particularly advisories from the U.S. Surgeon General, to set expectations for safety-by-design, transparency, and evaluation across platforms.

Each option has merits, but I recommend adopting Option A as the primary approach, supplemented by elements of B and C. Standards-based regulation best balances innovation with safety, ensuring that tools already in widespread use can be made safer rather than pushed underground. However, federal guidance helps coordinate national norms, while targeted bans remain appropriate for the highest-risk applications. Together, these policies form a layered response that prioritizes both youth protection and equitable access.

Problem Statement

Youth mental health needs are profound and growing, with nearly 2.8 million adolescents experiencing major depressive episodes in 2024 and close to 3 million reporting suicidal thoughts (Reinert et al., 2025). Shortages of school counselors, sometimes exceeding 700 students per one counselor (Inseparable, 2025), leave many without access to timely care. In this gap, AI tools have proliferated: predictive algorithms flagging suicide risk, and chatbots offering cognitive-behavioral support (Sharma et al., 2025; Li et al., 2025). While promising, these tools also pose risks, from unsafe or misleading responses (Harwell, 2024) to potential "AI psychosis" (Cha, 2025). The challenge is how to govern AI so it supports youth safely and equitably.

Policy Option 1: Safety Standards for AI Companions

One path forward is to regulate the design and operation of AI chatbots that interact with minors. California's Assembly Bill 1064, the *Leading Ethical AI Development for Kids Act*, offers a model. The bill requires AI companion tools marketed to children to include safeguards such as crisis-response protocols that escalate to human hotlines, clear disclosures that the chatbot is not human, and ongoing reporting on user impacts (California Legislative Information, 2025). This approach is backed by state legislators and child safety advocates who argue that regulation can preserve the accessibility and convenience of AI tools while reducing risks. The government tool here is regulation and standards-setting, combined with disclosure requirements to ensure transparency. In practice, this would mean that companies cannot market AI companions to youth unless they demonstrate compliance with baseline safety and accountability features.

Policy Option 2: State Bans and Restrictions on AI Therapy

A more restrictive approach has emerged in states such as Illinois and Nevada, which have enacted bans on the use of AI for therapeutic purposes in mental health. Utah has taken a middle path, permitting limited use of AI in mental health contexts but requiring stringent disclosures and strong privacy protections (Schoenherr, 2025). These proposals, advanced primarily by state legislators and attorneys general, are motivated by precaution. Critics of AI therapy argue that unproven tools should not substitute for human clinicians, especially when adolescents are at risk of crisis. The governmental tool here is command-and-control regulation in the form of outright prohibition, with mandatory disclosure layered on where use is allowed. This approach prioritizes minimizing harm even at the cost of slowing adoption of AI-based supports in schools and clinics.

Policy Option 3: Federal Public Health Guidance

At the federal level, the U.S. Surgeon General has issued advisories addressing youth mental health and the risks of digital platforms, including those driven by AI (Office of the U.S. Surgeon General, 2023). While these advisories do not carry the force of law, they represent a powerful form of public information and education, or what Kraft and Furlong (2021) describe as "soft law." By urging platforms to adopt safety-by-design principles, calling for transparency, and emphasizing independent evaluation, the Surgeon General sets a national benchmark for best practices. Federal agencies can then integrate these benchmarks into grant requirements, procurement policies, and public-health campaigns. This approach leverages the moral authority of federal health officials and creates consistency across jurisdictions, even in the absence of binding regulation.

Policy Recommendation

Among the three options, Option A, safety standards for AI companions, emerges as the most balanced and pragmatic policy choice. Regulation through design requirements addresses the core risks of unsafe chatbot behavior while allowing youth continued access to the supports many are already using. This approach aligns with evidence that regulation and disclosure can improve the safety of consumer technologies without stifling innovation (Sharma et al., 2025).

Option C, federal public health guidance, remains an essential complement, even though advisories lack enforcement power. Federal advisories play an important agenda-setting role,

shaping the expectations of school districts, community/private clinics, technology vendors, and parents. They also provide a framework that state policymakers and agencies can draw upon when crafting contracts, grant programs, or local rules.

Option B, state bans and restrictions, should serve as a tertiary tool, reserved for contexts where the risks of AI-delivered therapy clearly outweigh potential benefits. Bans are politically salient and can prevent immediate harms, but they also risk pushing vulnerable youth toward unregulated online tools. For this reason, bans should be narrowly tailored, ideally paired with exemptions for research and pilot programs that generate evidence for future policymaking.

Taken together, the recommended policy ranking is: Option A first, as the foundation for safe adoption; Option C second, to ensure national consistency and public communication; and Option B third, as a precautionary fallback where risks remain high. This ordering reflects the best balance of feasibility, cost-effectiveness, and equity, giving decision makers a clear path forward to protect young people while fostering responsible innovation.

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