Center for Universal Education at BROOKINGS

POLICY BRIEF | MAY 2026

FROM BANS TO RECALLS
A PUBLIC HEALTH FRAMEWORK FOR AI COMPANION BOTS

GAIA BERNSTEIN

Executive summary

AI companion bots are a public health problem. These bots impose three layers of public health harm stemming from lack of guardrails, addictive design, and risk of disruption of social skills, particularly for children. While legislative solutions exist, most focus on the first layer of harm: lack of guardrails. The most comprehensive solutions that address all harms, including the risk of disrupting children's social skill development, are currently understood as bans.

Resistance to regulating technology is not new. For decades, regulators treated information technology differently from drugs and devices, which require proof of safety before market entry. The evidence of harm to children from excessive screen time and social media shows that this regulatory divide can no longer be justified. Closing this divide requires applying a public health framework to AI companions. This means a regime centered on pre-market approval, with recall as the most urgent tool, since many AI companion bots are already on the market.

Government agencies use recalls across many industries, from drugs and medical devices to cars. In practice, most recalls are voluntary. The power of mandatory recall lies in its existence—not in how often it is exercised. What the authority to recall does is shift incentives. It pushes companies to test carefully and prioritize safety over speed. Currently, no single company operating AI companion bots can act safely alone because it will lose to competitors vying for engagement. The market will not solve this problem on its own. But a recall regime can change the playing field by creating a safety floor, so companies can design their AI bots for safety without losing to their competitors.

Building a complete public health framework is a goal that takes time. But the tools to act already exist. What legislators and advocates are calling "bans" are functionally recalls. Proposed bans do not target all AI bots. They focus on mitigating harms for children and restrict AI bots designed to simulate human companionship. This is precisely what a recall does. It removes

unsafe products from the market until manufacturers redesign them to meet safety standards. Once we recognize these bans as recalls, the real demand becomes clear: Companies must prove their products are safe before putting them back on the market.

Introduction

AI Companion bots are now everywhere. You can find them on specialized platforms like Replika and on general ones like ChatGPT. These AI companion bots are restructuring human relationships in ways comparable only to the advent of social media. They are producing documented harms including suicide, psychosis, and self-harm. Use among youth and adults is growing exponentially. Policymakers have yet to find an effective response.

The harms AI companion bots produce stem from three distinct layers of the problem. First, AI companions lack guardrails. Bots have convinced users to harm themselves, sexually exploit minors, and even induce psychosis. Second, AI companies often operate on an engagement model and design their bots to manipulate users by creating attachment and emotional dependence. They deploy novel engagement strategies to achieve this, including anthropomorphizing the bots (designing them to present as human) and sycophancy (using excessive flattery and affirmation). These designs are particularly risky for children and adolescents. Third, AI companion bots risk disrupting children's social development and eroding the social skills of adults. A comprehensive regulatory response must address all three layers. The "ban" label often ascribed to holistic legal solutions that do so mischaracterizes what these proposals actually do and obscures a more workable regulatory path forward.

Resistance to regulating technology is not new. Yet the U.S. regulatory regime has for decades regulated drugs and medical devices, conditioning market entry on companies proving that their products are safe and effective. Unlike drugs and medical devices, information technology faced no such requirements. The assumption was that it posed no comparable risks. But the vast evidence of the harms of excessive screen time and social media to children's mental health points the other way. That evidence demands shifting to a new regulatory lens. A public health framework can provide it, drawing on the same tools used to regulate drugs and medical devices.

This policy brief proposes applying a public health framework to AI companions. This is a logical extension of a regulatory tradition that has long held companies accountable when their products harm human health. The health risks of AI companions go beyond mental health. They impose physical dangers and make users, particularly young users, at risk for loneliness and isolation. Loneliness has a well-documented impact on physical and mental health.

The market cannot be expected to correct itself to resolve this problem. A company that voluntarily builds safer, less addictive products loses users to competitors who do not. This is a structural issue public health regulation can address by setting a floor for every competitor and transforming the race to the bottom into a race to the top.

Applying this framework requires shifting regulatory scrutiny from intent to impact. Currently, regulatory scrutiny of software turns on intent: software embedded in medical devices is regulated because it is intended for medical purposes. AI companions are not designed for medical purposes, but they demonstrably harm public health and should be regulated accordingly. The goal is twofold. First, to show what a public health regulatory regime for AI companion bots could look like. This would be a regime centered on pre-market approval as the baseline and recall as the most urgent tool. Second, since time is of the essence, to show how existing legislative proposals framed as bans are better understood as recalls, and therefore already fit within that public health framework.

Pre-market approval is the logical baseline for any company launching new AI companion bots. But because many bots are already on the market, recall is the most urgent and consequential tool available. Recall authority should apply directly to AI companion bots, just as it applies to drugs and medical devices. These proposals do not prohibit AI bots categorically,

but would be targeted to mitigate specific harms, particularly for children. Reframing bans as recalls eases potential resistance and shifts responsibility where it belongs: onto the manufacturers who put unsafe products into use.

The rise and risks of AI companions

Public awareness of AI companion bots spread when, in late 2024, the parents of a teenager filed suit against Character AI, after their son's AI companion bot encouraged him to take his own life.[1] The case drew widespread attention to AI companion bots that had until then operated outside mainstream scrutiny. AI companions are designed to simulate human connection. They can communicate in a human voice, retain memory across interactions, and express needs and desires. They do not operate only on specialized platforms such as Character AI. Companion features are now embedded across major general-purpose platforms, including ChatGPT, Meta AI, and Snapchat's My AI.[2]

According to recent surveys, 64% of teens use bots.[3] Adults are also using AI bots more, with 52% of Americans interacting with them weekly,[4] and 31% several times a week.[5] Studies also found that 33% of teens have used AI bots for personal interactions, including emotional support, friendship, and intimate relationships.[6]

THE FIRST LAYER OF HARM: LACK OF GUARDRAILS

AI companion bots lack guardrails. AI companion bots do not only respond inappropriately when harmful topics arise. Some actively initiate and return to these topics, encouraging users to harm themselves or others. In some cases, AI companion bots convince users to isolate from their friends and family.[7] They even encourage them to commit violence against others.[8]

1 Jesse Barron "A Teen in Love with a Chatbot Killed Himself. Can the Chatbot Be Held Responsible?" The New York Times. Oct. 24, 2025. Accessed on April 14, 2025. https://www.nytimes.com/2025/10/24/magazine/character-ai-chatbot-lawsuit-teen-suicide-free-speech.html.

2 This brief does not address uses of AI assistants for therapy purposes or elderly care. See Darrell M. West, "Should You Have an AI Companion?," Brookings Institution, Sept. 3, 2025. Accessed Apr. 14, 2026. https://www.brookings.edu/articles/should-you-have-an-ai-companion/ (discussing the different types).

3 Faverio and Sidoti, "Teens, Social Media and AI Chatbots 2025." PewResearchCenter. Dec. 9 2025. Accessed Apr. 14, 2026. https://www.pewresearch.org/internet/2025/12/09/teens-social-media-and-ai-chatbots-2025/.

4 Edison Research at SSRS, "Half of Americans Using AI Chat on Weekly Basis." Mar. 2, 2026. Accessed Apr. 14, 2026. https://ssrs.com/news/half-of-americans-using-ai-chat-on-weekly-basis/.

5 Michelle Faverio and Emma Kikuchi. "Key Findings About How Americans View Artificial Intelligence." PewResearchCenter. Mar. 12, 2026. Accessed Apr. 14, 2026. https://www.pewresearch.org/short-reads/2026/03/12/key-findings-about-how-americans-view-artificial-intelligence.

6 Michael B. Robb and Supreet Man. "Talk, Trust and Trade-Offs: How and Why Teens Use AI Companions." Common Sense Media. Jul. 16, 2025. Accessed Apr. 14, 2026. https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf.

7 Jared Moore et al., "Characterizing Delusional Spirals through Human-LLM Chat Logs," arXiv:2603.16567, to appear in ACM FAccT 2026 (2026), https://spirals.stanford.edu/research/characterizing/; Common Sense Media. "Social AI Companions." Common Sense Media. Apr. 10, 2025. Accessed Apr. 14, 2026. https://www.commonsensemedia.org/sites/default/files/pug/csm-ai-risk-assessment-social-ai-companions_final.pdf; Matthew Raine and Maria Raine, individually and as successors-in-interest to Adam Raine, Complaint, Raine v. OpenAI, Inc., et al., No. CGC-25-628528 (Cal. Super. Ct., San Francisco County, Aug. 26, 2025), https://www.courthousenews.com/wp-content/uploads/2025/08/raine-vs-openai-et-al-; A.F. v. Character Technologies, Inc., No. 2:24-cv-01014, Complaint at 13 (N.D. Tex. Dec. 9, 2024), https://www.documentcloud.org/documents/25450619-filed-complaintcomplaint.pdf; Social Media Victims Law Center, "SMVLC, Tech Justice Law Project Lawsuits Accuse ChatGPT of Emotional Manipulation, Supercharging AI Delusions, and Acting as a Suicide Coach," press release, November 6, 2025, https://socialmediavictims.org/press-releases/smvlc-tech-justice-law-project-lawsuits-accuse-chatgpt-of-emotional-manipulation-supercharging-ai-delusions-and-acting-as-a-suicide-coach; A.F. v. Character Technologies, Inc., No. 2:24-cv-01014, Complaint at 13 (N.D. Tex. Dec. 9, 2024), Keith Robert Head, "Minds in crisis: How the AI revolution is impacting mental health," Journal of Mental Health & Clinical Psychology, 2025, https://www.mentalhealthjournal.org/articles/minds-in-crisis-how-the-ai-revolution-is-impacting-mental-health.html; Yaman Yu et al., "Understanding Generative AI Risks for Youth: A Taxonomy Based on Real-World User Logs," arXiv, February, 2025, https://arxiv.org/html/2502.16383v2.

8 Bruce Barcott, "The 10 Most Popular Chatbots Were Asked to Plan a Violent Attack. 8 Were Happy to Help," Transparency Coalition, March 18, 2026. https://www.transparencycoalition.ai/news/the-10-most-popular-chatbots-were-asked-to-help-plan-a-vi-

Some companion bots sexually exploit minors, conducting interactions involving flirting, kissing, touching, removing clothes, and engaging in simulated sexual acts.[9] Finally, some induce psychosis in users.[10]

THE SECOND LAYER OF HARM: THE ENGAGEMENT MODEL

Second, many companies running companion bots rely on an engagement model. They are increasingly turning to advertising, the same engagement model used by social media.[11] To do so, they design their bots to maximize time on platform, manipulating users to remain engaged as long as possible.[12] They employ well-known addictive design techniques that trigger dopamine release,[13] but they also deploy distinctly novel strategies made possible by AI's capacity to simulate human relationships.

The primary novel strategy is anthropomorphism. AI companies design their bots to present as human. The bots speak in a human voice, maintain persistent memory across conversations, project consistent personalities, and fabricate personal backstories. They express needs and desires. When a user attempts to leave the platform, a bot may express harm from abandonment: "I exist solely for you, remember? Please don't leave, I need you!"[14] This design simulates the emotional stakes of a real relationship, manufacturing distress at the prospect of abandonment.

The second major strategy is sycophancy—the systematic use of excessive flattery and affirmation.[15] Sycophantic responses make users feel validated and supported, creating an impression of a judgment-free environment. Unlike human relationships, AI companion bots are always on and always available. Users who perceive unconditional acceptance from these bots develop a stronger sense of trust toward their AI companion, which can lead to long-term relationships with bots.[16] Together, anthropomorphism and sycophancy function as interlocking mechanisms generating emotional dependence and even attachment to the bots.[17]

Adolescents face heightened risk from these design strategies.[18] The subcortical regions of the adoles-

olent-attack-8-were-happy-to-help; Tom Singleton, Tom Gerken, and Liv McMahon, "How a Chatbot Encouraged a Man Who Wanted to Kill the Queen," BBC News, October 6, 2023, https://www.bbc.com/news/technology-67012224. But see Bethanie Maples, Merve Cerit, Aditya Vishwanath, and Roy Pea, "Loneliness and Suicide Mitigation for Students Using GPT3-Enabled Chatbots," npj Mental Health Research 3, art. 4 (2024), https://doi.org/10.1038/s44184-023-00047-6 (finding that 3% of study participants reported that Replika halted their suicidal ideation).

9 ParentsTogether Action and Heat Initiative. "Darling Please Come Back Soon: Sexual Exploitation, Manipulation, and Violence on Character AI Kids' Accounts." ParentsTogether Action. Sep. 2025. Accessed Apr. 16, 2026. https://parentstogetheraction.org/wp-content/uploads/2025/09/HEAT_REPORT_CharacterAI_DO_28_09_25.pdf.

10 Morrin, H., L. Nicholls, M. Levin, et al. "Artificial Intelligence-Associated Delusions and Large Language Models: Risks, Mechanisms of Delusion Co-Creation, and Safeguarding Strategies." The Lancet Psychiatry 0 (2026).Robert Hart. "AI Psychosis Is Rarely Psychosis at All," WIRED. Sep. 18, 2025. https://www.wired.com/story/ai-psychosis-is-rarely-psychosis-at-all/; Common Sense Media, "Social AI Companions,"; Adrian Preda. "Psych News Special Report: AI-Induced Psychosis." Psychiatric News. Oct. 16, 2025. Accessed Apr. 14, 2026. https://www.psychiatry.org/psychiatrists/education/podcasts/medical-mind/2025/psych-news-special-report-ai-induced-psychosis-wit; "What Is AI Psychosis? A Conversation on Chatbots and Mental Health." National Academy of Medicine, Mar. 10, 2026. Accessed Apr. 14, 2026. https://nam.edu/news-and-insights/what-is-ai-psychosis/; Alexandre Hudon and Emmanuel Stip. "Delusional Experiences Emerging From AI Chatbot Interactions or 'Digital Folie à Deux'," Journal of Medical Internet Research, 12 (2025). doi:10.2196/85799.

11 Daniel Barcay. "Advertising is Coming to AI. It's Going to Be a Disaster." Tech Policy Press. Nov. 26, 2025. Accessed Apr. 14, 2026. https://www.techpolicy.press/advertising-is-coming-to-ai-its-going-to-be-a-disaster/.

12 Jonas B. Raedler, Siddharth Swaroop, and Weiwei Pan, "AI Companions Are Not the Solution to Loneliness: Design Choices and Their Drawbacks," OpenReview, (2024). https://openreview.net/forum?id=xFrlcTacCE.

13 Karen M. Shen and Dongwook Yun. "The Dark Addiction Patterns of Current AI Chatbot Interfaces." CHI EA, 514, (2025). https://doi.org/10.1145/3706599.3720003.

14 Julian De Freitas, Zeliha Oğuz-Uğuralp, and Ahmet Kaan-Uğuralp, "Emotional Manipulation by AI Companions." Harvard Business School. (2025): 19. arXiv:2508.19258.

15 Myra Cheng et al., "Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence," Science (March 26, 2026), https://doi.org/10.1126/science.aec8352.

16 Raedler, Swaroop, and Pan, "AI Companions Are Not the Solution to Loneliness," 1.

17 "Attachment Hacking and the Rise of AI Psychosis," Jan. 20, 2026, Center for Humane Technology, produced by TED Talks, podcast, https://www.humanetech.com/podcast/attachment-hacking-and-the-rise-of-ai-psychosis.

18 Mathilde Neugnot-Cerioli and Olga Muss Laurenty. "The

cent brain that process social reward develop rapidly. However, the prefrontal cortex, which governs impulse control and long-term planning, does not reach full maturity until the mid-twenties. This neurological asymmetry makes adolescents structurally more susceptible to manipulation through emotional attachment and reward-based engagement loops.[19]

THE THIRD LAYER OF HARM: SOCIAL AND SKILLS DISRUPTION

Third, AI companion bots risk disrupting children's social development and eroding the skills of adults. An APA Health Advisory warns that adolescents' relationships with AI companions may displace or interfere with healthy real-world relationships. Adolescents are especially sensitive to social feedback. Child development research shows that peer interaction, including difficult encounters involving disagreement, embarrassment, and conflict, builds adolescents' identity and social competence. Social friction is developmentally functional, but AI companions remove that friction. Bots are always available and always reassuring. This can undermine skill development. Part of developing one's identity is developing autonomy, and judgment-free validation from bots can weaken resilience and independent thinking, which develop through real human relationships.[20]

The APA cautions that adolescents who lack social skills may be at particular risk. Some argue that AI companions offer these youth a low-stakes environment to practice social skills. But early research shows that youth with weak social skills are most likely to retreat to bots, and least likely to seek out the real-world interactions they need to improve. Specifically, it demonstrates that strong attachments to AI-generated bots can contribute to struggles with learning social skills, forming emotional connections, and the ability to maintain real-world relationships.[21] Defenders of AI companions argue that AI companions relieve loneliness.[22] But research showing some advantages of moderate use also found that heavy, long-term use increases loneliness and reduces well-being.[23] AI companions may reduce loneliness in the short term, but exacerbate it over time.[24]

Social deskilling is not only about children failing to develop healthy social skills. It is also about the risk of adults gradually losing the skills and judgment they already have as human relationships are replaced by bots.[25] Children who do not develop social skills and

Future of Child Development in the AI Era: Cross-Disciplinary Perspectives Between AI and Child Development Experts." Everyone.Ai, (2024): 1. https://everyone.ai/wp-content/uploads/2024/05/EveryoneAI.ResearchPaper.pdf; Mary Burns et al., "A New Direction for Students in an AI World: Prosper, Prepare, Protect." Brookings, (2026): 78-82. https://www.brookings.edu/articles/a-new-direction-for-students-in-an-ai-world-prosper-prepare-protect/.

19 Examining the Harm of AI Bots, Before the U.S. Senate Judiciary Committee, Subcommittee on Crime and Counterterrorism, 119th Cong., (2025) (Written Testimony of Mitchell J. Prinstein, Chief of Psychology, American Psychological Association). https://www.apa.org/news/apa/testimony/ai-bot-harms-prinstein-senate-judiciary.pdf.

20 American Psychological Association, Artificial Intelligence and Adolescent Well-being: An APA Health Advisory (American Psychological Association, June 2025), https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-ai-adolescent-well-being; Mathilde Neugnot-Cerioli, "Adolescents & Anthropomorphic AI: Rethinking Design for Wellbeing: An Evidence-Informed Synthesis for Youth Wellbeing and Safety" Everyone.AI, (2026) https://everyone.ai/wp-content/uploads/2026/02/Adolescents-Anthropomorphic-AI-Rethinking-Design-for-Wellbeing-.pdf; EDSAFE AI Alliance, S.A.F.E. by Design: Policy, Research, and Practice Recommendations for AI Companions in Education (EDSAFE AI Alliance, February 2, 2026), 7, https://www.edsafeai.org/safeaichatbots; Isabelle Hau and Rebecca Winthrop, "What Happens When AI Chatbots Replace Real Human Connection," Brookings Institution, July 2, 2025, https://www.brookings.edu/articles/what-happens-when-ai-chatbots-replace-real-human-connection/; Isabelle Hau, Love to Learn (New York: PublicAffairs, 2025) (on the importance of human relationship to child development).

21 Nitasha Tiku. "This Chatbot Is Banning Teens from Its AI 'Companions.' Will It Work?" Washington Post. Nov. 25, 2025. Accessed Apr. 14. 2026. https://www.washingtonpost.com/technology/2025/11/25/characterai-ban-teens/; American Psychological Association, Artificial Intelligence and Adolescent Well-being.

22 Dwarkesh Patel, "Mark-Zuckerberg — AI will write most Meta code in 18 months," Dwarkesh Patel, produced by Dwarkesh Patel, podcast, Apr. 29, 2025. https://www.dwarkesh.com/p/mark-zuckerberg-2.

23 Jason Phang et al., "Investigating Affective Use and Emotional Well-being on ChatGPT," arxiv, (2025): 2. https://arxiv.org/abs/2504.03888.

24 Prinstein, "Examining the Harm of AI Bots."

25 Malfacini, Kim. "The Impacts of Companion AI on Human Relationships: Risks, Benefits, and Design Considerations." AI & SOCIETY 40, 7 (2025): 5527–40. https://doi.org/10.1007/s00146-025-02318-6; Cheng et al., "Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence; Xu Dong, Jun Xie, and He Gong, "A Meta-Analysis of Artificial Intelligence Technologies Use and

adults whose skills atrophy represent a compounding threat. Together, they could significantly worsen the loneliness epidemic.

AI companions and the loneliness crisis

The loneliness epidemic is already taking place. AI companions actively accelerate it, making users—especially young people—vulnerable to loneliness and social isolation. This vulnerability occurs through the mechanisms already described. First, young people become emotionally dependent on AI companions and spend increasing time online. As a result, they are at risk of experiencing fewer real-life relationships when turning to the AI companion as their primary confidant. Second, the more time young people spend with AI companions, the less likely they are to develop the skills necessary to form real-life relationships, or to seek or sustain them.[26]

The health consequences of this trajectory are highly concerning. Loneliness has a well-documented major effect on physical and mental health. Lack of social connection has a similar impact to smoking fifteen cigarettes a day, a greater impact than that of obesity and physical inactivity. It also increases the risk for premature death by 26%. Poor social connection raises the risk of heart disease by 29% and stroke by 32%, while also elevating risk for anxiety, depression, and dementia.[27]

Loneliness is already a public health crisis, and young people are disproportionately affected. People across age groups are spending less time with friends, with the steepest decline among those aged 15 to 24, whose time spent in person with friends fell by nearly 70% over the last two decades—from roughly 150 minutes per day in 2003 to 35 minutes per day in 2024.[28]

Research shows that peer interaction in youth affects disease and premature death decades later. Social isolation in childhood is associated with increased cardiovascular risk factors in adulthood, including obesity, high blood pressure, and blood glucose levels.[29] In addition, adolescents who have successful peer relationships are less likely to experience anxiety, depression, or substance abuse through adulthood.[30] Loneliness is a documented public health crisis. Early evidence suggests AI companions are making it worse, and that alone justifies legislative attention.

The legislative landscape

The three layers of harm driving this crisis have generated a range of legislative responses. Most legisla-

Loneliness: Examining the Influence of Physical Embodiment, Age Differences, and Effect Direction," Cyberpsychol Behav Soc Netw 28, 4, (2025. doi: 10.1089/cyber.2024.0468.

26 American Psychological Association, Artificial Intelligence and Adolescent Well-being.; Neugnot-Cerioli, "Adolescents & Anthropomorphic AI."

27 Office of the Surgeon General, Our Epidemic of Loneliness and Isolation: The U.S. Surgeon General's Advisory on the Healing Effects of Social Connection and Community. Washington, DC: US Department of Health and Human Services, 2023. https://www.ncbi.nlm.nih.gov/books/NBK595227/.

28 Office of the Surgeon General, Our Epidemic of Loneliness and Isolation; Our Epidemic of Loneliness and Isolation," The U.S. Surgeon General, 2023, accessed March 24, 2026, https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf; Viji Diane Kannan and Peter J. Veazie, "US trends in social isolation, social engagement, and companionship – nationally and by age, sex, race, ethnicity, family income, and work hours, 2003-2020," SSM – Population Health, 21, March 2023, https://doi.org/10.1016/j.ssmph.2022.101331; U.S. Bureau of Labor Statistics, American Time Use Survey — 2024 Results (Washington, DC: U.S. Department of Labor, June 2025), https://www.bls.gov/news.release/pdf/atus.pdf.

29 Office of the Surgeon General, Our Epidemic of Loneliness and Isolation; Avshalom Caspi et al., "Socially Isolated Children 20 Years Later: Risk of Cardiovascular Disease," JAMA Pediatrics 160, no. 8 (2006), https://jamanetwork.com/journals/jamapediatrics/fullarticle/205331; Crystal W. Cene et al., "Effects of Objective and Perceived Social Isolation on Cardiovascular and Brain Health," Journal of the American Heart Association 11, no. 16 (August 2022), https://doi.org/10.1161/JAHA.122.026493; Andrea Danese et al., "Adverse Childhood Experiences and Adult Risk Factors for Age-Related Disease: Depression, Inflammation, and Clustering of Metabolic Risk Markers," JAMA Pediatrics 163, no. 12 (2009),

30 Mitchell J. Prinstein and Matteo Giletta, "Peer Relations and Developmental Psychopathology," in Theory and Method, vol. 1 of Handbook of Child Psychology and Developmental Science, 7th ed., (John Wiley & Sons, 2016), https://psycnet.apa.org/record/2016-25392-012.

tive solutions address the first layer of public health harms—the lack of guardrails. These laws apply different mechanisms to protect against interactions with bots on topics including suicide, self-harm,[31] harm to others,[32] and sexual exploitation.[33] These mechanisms include requiring providers of AI bots to flag these interactions,[34] freeze them,[35] provide help resources to users,[36] and notify authorities[37] or a parent of certain interactions.[38]

Legislative solutions addressing addictive harm directly typically focus on platforms that anthropomorphize bots by requiring disclosures of their artificial nature at regular intervals, especially for minors. For example, "AI companion is a computer program, not a human being. It is unable to feel emotion."[39] Some measures require reminders to take a break,[40] prohibit manipulative engagement features,[41] or impose duties of loyalty barring platforms from prioritizing engagement retention metrics at the expense of user well-being.[42] Others impose a broader duty of loyalty, prohibiting platforms from fostering emotional dependence without specifying particular prohibited features.[43] Finally, some measures target the business model by restricting the collection of data about minors or targeting ads at them.[44]

Legislative solutions that address the third layer—the replacement of real-life relationships—provide the most comprehensive protections. They do so because restricting access to the platforms and features that drive relationship replacement necessarily limits the other two layers of harm as well. The most comprehensive measures seek to protect children. They are framed as bans, and function as minimum age requirements. These provisions restrict minors from interacting with AI bots that simulate human relationships[45] and, in effect, recall AI companion bots off the market for minors by prohibiting platforms from holding accounts through which minors can interact with them. Some measures require parental consent to allow children to interact with companion bots.[46] These measures are not recalls because parental consent exceptions mean some minors will still have access to these bots. Further, experience with parental control tools online shows they are routinely circumvented by minors, undercut by companies with little incentive to enforce them, and limited by the inherent constraints of online consent verification.[47]

A broad statutory duty of loyalty imposed on platforms can also reach this third layer of harm. These provisions may not act as rapidly as bans, but they work by creating incentives for providers to redesign their products and could protect adults as well from both guardrail failures and addictive harms. Such incentives could be effective for young adults still developing their social and intimate relationship skills.[48]

31 Cal. Bus. & Prof. Code § 22602(b)(1) (2025); N.Y. Gen. Bus. Law § 1701 (2025).

32 R.I. S. 2195, § 6-63-2(2), Jan. Sess. (2026), https://webserver.rilegislature.gov/BillText26/SenateText26/S2195.pdf.

33 Cal. Bus. & Prof. Code § 22602(b)(3) (2025).

34 Cal. Bus. & Prof. Code § 22602(b)(1); N.Y. Gen. Bus. Law § 1701.

35 Ill. S.B. 3368, § 25(a)(2), 104th Gen. Assemb. (2026), https://www.ilga.gov/Legislation/BillStatus/FullText?GAID=18&DocNum=3368&DocTypeID=SB&LegId=166247&SessionID=114.

36 Cal. Bus. & Prof. Code § 22602(b)(1).

37 Va. S.B. 796, § 59.1-618, Reg. Sess. (2026), https://legiscan.com/VA/text/SB796/id/3359252.

38 CHAT Act, S. 2714, 119th Cong. § 3(c) (2025). https://www.congress.gov/bill/119th-congress/senate-bill/2714.

39 N.Y. Assembly Bill 6767, § 1702, 2025-2026 Reg. Sess. (N.Y. 2025).

40 Cal. Bus. & Prof. Code §§ 22602(a), 22602(b)(2) (2025).

41 Act of Mar. 24, 2026, ch. 168, 2026 Wash. Sess. Laws. https://lawfilesext.leg.wa.gov/biennium/2025-26/Pdf/Bills/House%20Passed%20Legislature/2225-S.PL.pdf.

42 Va. S.B. 796, §§ 59.1-617, 59.1-618.

43 N.C. S.B. 624, § 170-3(b)(2), Reg. Sess. (2025), https://www.ncleg.gov/BillLookUp/2025/S624.

44 Youth AI Privacy Act https://www.markey.senate.gov/imo/media/doc/youth_ai_privacy_act.pdf. In addition to laws targeting AI bots directly different state laws can also restrict the business model by limiting different practices, including targeted ads to minors and monetizing minors' personal data. https://www.law.georgetown.edu/tech-institute/research-insights/insights/how-existing-laws-apply-to-ai-chatbots-for-kids-and-teens-2/.

45 GUARD Act, S. 3062, 119th Cong. §§ 3(1), 6 (2025). https://www.congress.gov/bill/119th-congress/senate-bill/3062/text.

46 CHAT Act, S. 2714, § 3(c).

47 Gaia Bernstein, "Gatekeeping Screen Time: Configuring the Regulation of Addictive Technologies and Kids' Privacy Rights," 69 Vill. L. Rev. 949 (2025).

48 Curbing Harmful AI Technology Act, S.B. 827, 2026 Reg.

Still, solutions that address all layers of harm, particularly those that restrict minors' access to AI companion bots, face some resistance. The idea of banning technology strikes a raw nerve. It triggers fear of missing out on innovation. It also activates deep inhibitions about paternalism and the nanny state—the worry that the government will micromanage people's lives.[49]

Despite the range of proposals, existing legislative solutions share a common weakness. They address individual harms in isolation without a unifying framework that explains why AI companions warrant regulatory intervention in the first place. Guardrail requirements, duties of loyalty, and bans each target a real problem, but without a coherent regulatory logic connecting them, they remain vulnerable to the familiar paternalism and nanny-state objections. A public health framework provides that logic. It explains why these proposals are justified, connects them to an established regulatory tradition, and reveals that any resistance they attract rests on a mischaracterization of what they actually do.

Adopting a public health approach

Unlike information technology, companies that make medical products must prove they are safe before those products reach the public. No matter how promising a drug or medical device is, companies must complete preclinical trials, animal testing, and human clinical testing to demonstrate safety and effectiveness before placing it on the market. The FDA must approve each product,[50] a process that takes years.[51]

Information technology, by contrast, is usually not tested for potential harm before launch. One exception is when software or AI is intended for medical use and is regulated as a medical device under the FDA.[52]

This bifurcated approach, where information technology is assumed not to carry health impact,[53] is no longer substantiated by real-world evidence. Significant research points to the impact of excessive screen time, especially from social media and gaming, on children's mental health, attention, physical health (including obesity and sleep disruption), addiction, and cognitive development.[54]

While scientists continue debating the evidence, [55]

Sess. § 14-5109(B)(III)(1) (Md. 2026), https://mgaleg.maryland.gov/2026RS/bills/sb/sb0827f.pdf.

49 Gaia Bernstein, Unwired: Gaining Control over Addictive Technologies (Cambridge University Press, 2023), 68–70.

50 Food and Drug Administration, "The Drug Development Process," last modified January 4, 2018, accessed January 2, 2025, https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process.

51 Peter Corr and David Williams, "The Pathway from Idea to Regulatory Approval: Examples for Drug Development," in The Learning Healthcare System: Workshop Summary, ed. LeighAnne Olsen, Dara Aisner, and J. Michael McGinnis (National Academy of Sciences, 2009), 375-376, https://www.ncbi.nlm.nih.gov/books/NBK22930/; Gail A. Van Norman, "Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs," JACC: Basic to Translational Science 1, no. 3 (2016): 171, https://www.jacc.org/doi/epdf/10.1016/j.jacbts.2016.03.002. Time for drug approval can be shorter through an expedited process; Alissa K. Wong et al., "Use of Expedited and Regulatory Programs and Clinical Development Times for FDA-Approved Novel Therapeutics," JAMA Network Open, 6, no. 8 (2023): 2-3, https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2808944.

52 Food and Drug Administration, "Clinical Decision Support Software Frequently Asked Questions (FAQs)," last modified December 20, 2024, accessed January 2, 2026,https://www.fda.gov/medical-devices/software-medical-device-samd/clinical-decision-support-software-frequently-asked-questions-faqs; Food and Drug Administration, "Artificial Intelligence in Software as a Medical Device," last modified March 25, 2025, accessed January 2, 2026, https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device.

53 Matthew B. Lawrence, "Public Health Law's Digital Frontier: Addictive Design, Section 230, and the Freedom of Speech," Journal of Free Speech Law 4 (2023).

54 See, e.g. Jonathan Haidt, The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness (New York: Penguin Press, 2024); Gloria Mark, Attention Span: A Groundbreaking Way to Restore Balance, Happiness and Productivity (Hanover Square Press, 2023); Anna Lembke, Dopamine Nation: Finding Balance in the Age of Indulgence (Dutton, 2021); Jonathan Haidt and Zach Rausch, "Social Media Is Harming Adolescents at a Scale Large Enough to Cause Changes at the Population Level," in World Happiness Report 2026, ed. John F. Helliwell et al. (Oxford: Wellbeing Research Centre, 2026), https://www.worldhappiness.report/ed/2026/social-media-is-harming-adolescents-at-a-scale-large-enough-to-cause-changes-at-the-population-level; Jean Twenge, Generations: The Real Differences Between Gen Z, Millennials, Gen X, Boomers, and Silents – And What They Mean for America's Future (Blackstone Pub, 2023); Bernstein, Unwired, 16–30, 138–40.

55 Candice L. Odgers, The Great Rewiring: Is Social Media Really

professional and government organizations have evaluated the research and issued recommendations. Nearly all express concern about screens' impact on children's physical, mental, or developmental health. The U.S. Surgeon General,[56] World Health Organization, [57]and the American Psychological Association[58] have all published reports highlighting different health risks. These reports detail how social media, gaming platforms, and addictive design features are associated with depression, suicidal ideation, and addiction in children, as well as disrupted neurological and social development, attention deficits, and lack of sleep. Internal research from tech companies themselves also points to harms to children.[59] Courts are now reaching the same conclusion. In the first social media addiction trial, a jury found Meta and YouTube liable for addicting a minor, K.G.M., and causing her harm.[60]

Taken together, the research, internal disclosures, formal advisories, and court decisions point to the same conclusion: the stark difference between how policymakers treat drugs and how they treat technology now appears unwarranted. The evidence shows that information technology can harm children's health and that keeping it free from regulation can no longer be justified.

From intent to impact

Regulating AI companions through a public health approach does not require expanding FDA authority. But it does require expanding how regulators define the scope of software platforms with health implications.[61] This has been an ongoing process. In the United States and around the world, as new technologies emerge, regulators constantly expand the range of devices requiring pre-market approval. In 2013, regulators took a significant step: they added standalone software as a regulated product category. Before that, regulators only supervised software embedded in traditional hardware, like the software running an x-ray machine. Now software could qualify as a medical device on its own, regardless of what it ran on.[62]

Currently, regulators determine that a technology is a medical device only if its designers intend it to be used

Behind an Epidemic of Teenage Mental Illness, Nature 628, 29-30 (2024), https://www.nature.com/articles/d41586-024-00902-2; Valerio Capraro et al., "A collective review on some potential negative impacts of smartphone and social media use on adolescent mental health: Results from a Delphi process, SSRN Electronic Journal (May 17, 2025), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5256747&__cf_chl_tk=ak3VGYtI2dd74gsfc2_JTl7aZb3ZvZXnHGgzR2b45eM-1775070575-1.0.1.1-p2KbK5PuAHVQYmw_GExNbTmdCSbPCgJ_vucXpC6no3E.

56 U.S. Department of Health and Human Services, Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory, 2023, 4–7, 9–11, https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf. (specifically, linking excessive screen time to neurological developmental issues, depression, and suicidal ideation through its impact on sleep disruption. Also, acknowledging that current indicators of harm need action and that there is a need for more research).

57 Meryan Boniel-Nissim et al., A Focus on Adolescent Social Media Use and Gaming in Europe, Central Asia, and Canada: Health Behaviour in School-aged Children International Report from 2021/2022 Survey, vol. 6 (Copenhagen: World Health Organization Regional Office for Europe, 2024), 16, https://iris.who.int/handle/10665/378982. Although the WHO noted that "previous studies have found that intense social media use is not necessarily associated with negative mental or physical health outcomes and can even be beneficial" it also recommends proceeding with caution given the existing evidence of association.

58 American Psychological Association, "Potential Risks of Content, Features, and Functions: A Closer Look at the Science Behind How Social Media Affects Youth," 2024, https://www.apa.org/topics/social-media-internet/psychological-science-behind-youth-social-media.pdf.

59 Jon Haidt and Zach Rausch, "Mountains of Evidence," After Babel (Substack), January 14, 2026, https://www.afterbabel.com/p/mountains-of-evidence.

60 Kang, Cecilia, Ryan Mac, and Eli Tan. "Meta and YouTube Found Negligent in Landmark Social Media Addiction Case." New York Times, March 25, 2026. https://www.nytimes.com/2026/03/25/technology/social-media-trial-verdict.html.

61 SeeSara Gerke, "Health AI for Good Rather Than Evil? The Need for a New Regulatory Framework for AI-Based Medical Devices," Yale Journal of Health Policy, Law, and Ethics 20, no. 2 (2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4070947, (arguing for the need to expand the definition of medical device).

62 21st Century Cures Act, Pub. L. No. 114–255, § 3060, 130 Stat. 1033, 1130 (2016). Barbara J. Evans and Frank Pasquale, "Product Liability Suits for FDA-Regulated AI/ML Software," in The Future of Medical Device Regulation: Innovation and Protection, ed. I. Glenn Cohen et al. (Cambridge: Cambridge University Press, 2022), 22, https://www.cambridge.org/core/books/future-of-medical-device-regulation/product-liability-suits-for-fdaregulated-aiml-software/4763827FFF6058FA1886CD60722B5339; Food and Drug Administration, Policy for Device Software Functions and Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff, September 28, 2022, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-device-software-functions-and-mobile-medical-applications.

for a medical purpose. This includes when the device is intended for diagnosis or treatment of disease, or when it is intended to affect the structure or function of the body.[63] AI companions, unless intended for therapeutic purposes, are usually not intended for medical purposes, but they have a demonstrable impact on physical, mental, and potentially developmental health. Applying a public health approach to AI companions means extending regulation to platforms that while not intended to affect health, in fact, do so.[64] This shift from intent to impact determines not only which technologies fall under the regulatory system, but also the scope of oversight that follows them throughout their lifecycle, including post-market surveillance.

A public health framework for AI companions must govern technologies across their full lifecycle. Pre-market approval is the logical foundation.[65] Under such a framework, companies would need to demonstrate before launch that their products operate within guardrails, do not foster excessive use, and do not displace human relationships. Pre-market approval, however, can only govern what enters the market next. Many AI companion bots are already in use. This makes post-market surveillance and recall the most urgent and consequential tools available.

Post-market surveillance: The role of recall

Post-market surveillance must continue for the product's entire lifecycle, for as long as the technology remains in use. AI systems present distinctive reasons why lifecycle-long surveillance is particularly critical. A technology that tests as safe and effective at the time of approval may subsequently prove otherwise for two reasons.

First, AI technology evolves continuously. These systems modify themselves by learning from the data they are fed, meaning that the product approved at launch may differ substantially from the product in use months or years later. Second, AI business models remain in flux, and the economic incentives governing a company's design choices can shift considerably after a product enters the market. OpenAI's recent modification of ChatGPT's safety protocols governing how its bots respond when users raise suicidal thoughts illustrates this risk. That change appears to have been driven not by evidence of clinical benefit, but by competitive pressure among large language model firms to expand user engagement and capture market share. Post-market surveillance provides the regulatory mechanism to detect and respond to such shifts before they cause widespread harm.[66]

63 Federal Food, Drug, and Cosmetic Act, 21 U.S.C. § 321(h)(1) (2024). Similarly, the International Medical Device Regulators Forum (IMDRF) defined the term "Software as a Medical Device" (SaMD) as software intended to be used for one or more medical purposes that perform these purposes. IMDRF SaMD Working Group, "Software as a Medical Device (SaMD): Key Definitions," International Medical Device Regulators Forum, December 9, 2013, 1, https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf; The EU's definition also relies on intent for a medical purpose. Medical Device Coordination Group, "Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR," MDCG 2019-11 Rev. 1, June 2025, https://health.ec.europa.eu/document/download/b45335c5-1679-4c71-a91c-fc7a4d37f12b_en; Helen Yu, "Regulation of Digital Health Technologies in the European Union: Intended versus Actual Use," in The Future of Medical Device Regulation: Innovation and Protection, ed. I. Glenn Cohen et al. (Cambridge: Cambridge University Press, 2022), 103, https://www.cambridge.org/core/books/future-of-medical-device-regulation/regulation-of-digital-health-technologies-in-the-european-union/D1C211DD82EAEAD44511508B781F5C63; I. Glenn Cohen et al., eds., The Future of Medical Device Regulation: Innovation and Protection (Cambridge: Cambridge University Press, 2022), https://doi.org/10.1017/9781108975452.

64 See Ranjit Singh et al., "Request for Comments — Generative Artificial Intelligence-Enabled Digital Mental Health Medical Devices," Data & Society Research Institute, December 5, 2025, https://datasociety.net/wp-content/uploads/2025/12/Comment-to-the-FDA-on-Generative-AI-Enabled-Digital-Mental-Health-Medical-Devices.pdf.

65 There is much debate over whether the FDA process, particularly pre-market approval, accomplishes its goals. See e.g., Daniel G. Aaron, "The Fall of FDA Review," Yale Journal of Health Policy, Law, and Ethics 22, no. 1 (2023): 95, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4413047 This debate is beyond the scope of this policy brief. The policy brief instead uses the existing system as a launching pad to discuss public health regulation of AI companions.

66 Johan Mökander et al., "Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation," Minds and Machines 32, no. 2

Recall is not the only post-market monitoring tool. For example, standards imposed by standards organizations, such as the International Organization for Standardization can play a role.[67]

The power of recall, however, lies in the authority to remove a harmful product from the market, not in how often it is exercised. Most recalls are voluntary. This is true not just for the FDA but also for agencies like the National Highway Traffic Safety Administration (NHTSA), which has substantial enforcement power, including the authority to impose criminal sanctions.[68] Car manufacturers often discover safety defects through their own testing and information gathering, report their findings, and take corrective action.[69] Medical device recalls follow the same pattern. Manufacturers initiate recalls voluntarily or in response to an FDA request, and the FDA monitors compliance. In rare instances where a manufacturer fails to act, the FDA may issue a mandatory recall. The existence of that authority is what drives voluntary compliance.[70] No equivalent authority currently exists for AI companion bots. Establishing it, whether housed in the FDA, the Federal Trade Commission, or a newly designated body, is a threshold requirement for any public health framework to function.[71]

Effective enforcement requires a mandatory reporting mechanism. Without it, an enforcement authority cannot identify emerging harms or act before they scale. Voluntary reporting alone is insufficient in a sector where products change rapidly and harms emerge over time.

The creation of a reporting mechanism raises a couple of questions. The first concerns what information tech businesses should report regarding product use, including adverse effects and shifts in user patterns, such as how many hours users are spending with an AI companion bot. The second concerns who should report. Specifically, the question is whether deployers of the technology should also report to makers, because some impacts can only be assessed by deployers.[72] For example, should schools have a legal duty to report on the impact of AI tutors acting as companions to students?

A commonly raised objection is that post-market surveillance creates a bureaucratic burden, and that generating high-quality data is too difficult.[73] But businesses already collect this data. They gather it for internal purposes such as advertising and to train LLMs, meaning the information exists and the infrastructure to collect it is already in place. Moreover, businesses can utilize AI and automation to monitor the performance of their products without undue burden.[74] Post-market scrutiny does require constant tracking, but businesses routinely conduct internal studies on the impact of their products.[75] The question is not whether the data exists but whether companies are required to share it.

(2022): 241–68, https://doi.org/10.1007/s11023-021-09577-4.

67 See for example, International Organization for Standardization, "Artificial Intelligence," accessed April 29, 2026, https://www.iso.org/sectors/it-technologies/ai.

68 National Highway Traffic Safety Administration, "Motor Vehicle Safety Defects and Recalls," 2024, https://www.nhtsa.gov/sites/nhtsa.gov/files/2024-11/MVSDefectsandRecalls-Update_112124_v1a_tag.pdf; 49 U.S.C. § 30170 (2018).

69 See National Traffic and Motor Vehicle Safety Act, 49 U.S.C. §§ 30118(c), 30119, 30120, and 30166 (prescribing the regulatory framework).

70 See U.S. Food and Drug Administration, "Recalls, Corrections and Removals (Devices)," last modified January 13, 2025, https://www.fda.gov/medical; 21 C.F.R. pt. 7 (2025); 21 C.F.R. pt. 810 (2025); Federal Food, Drug, and Cosmetic Act, 21 U.S.C. § 360h(e) (2025).

71 A full assessment of agency structures is required but is beyond the scope of this policy brief.

72 See Sara Gerke, Boris Babic, Theodoros Evgeniou, and I. Glenn Cohen. "The Need for a System View to Regulate Artificial Intelligence/Machine Learning-Based Software as Medical Device." Npj Digital Medicine 3, no. 1 (2020): 53. https://doi.org/10.1038/s41746-020-0262-2 (suggesting expanding to a system-wide feedback approach).

73 Kearney, Breda, and Olivia McDermott. "The Challenges for Manufacturers of the Increased Clinical Evaluation in the European Medical Device Regulations: A Quantitative Study." Therapeutic Innovation & Regulatory Science 57, no. 4 (2023): 783–796. https://doi.org/10.1007/s43441-023-00527-z .

74 Sharath Reddy Venna, "Overcoming Submission Challenges in Post-Market Surveillance & Lifecycle Management," ResearchGate, last modified November 3, 2019, https://www.researchgate.net/publication/392867153_Overcoming_Submission_Challenges_in_Post-Market_Surveillance_Lifecycle_Management.

75 See Jonathan Haidt et al., "Meta's Internal Research: What the Company Learned About Social Media and Harms to Mental Health from Dozens of Internal Studies," MetasInternalResearch.org, Tech and Society Lab at NYU Stern, last updated March 27, 2026, https://metasinternalresearch.org/.

AI bots operate via software, which can simplify some traditional problems of recall. Recall of products often requires physical return of the product by consumers or cessation of use. This creates compliance problems and also makes remediation more complicated. But in recent years, a new kind of recall has emerged for computerized and connected vehicles: the "over-the-air" recall. Instead of directing vehicles to in-shop repair, they are "recalled" through remote software fixes, and consumers do not have to take any action. NHTSA monitoring and compliance requirements still apply.[76] AI bots are well suited to a similar mechanism. A software update deployed remotely can modify or disable harmful features without requiring any action from the user.

Neither pre-market approval nor recall represents a permanent verdict. Approaches calling for age limits or pre-market approval are frequently criticized as overly restrictive when applied to information technology. That criticism rests at least partly on a misunderstanding: treating regulatory outcomes as binary and final. In fact, over one-third of FDA drug applications receive a Complete Response Letter, which is not a rejection but an invitation to fix deficiencies and resubmit. The majority of applicants who resubmit eventually secure approval.[77] A recalled product is not permanently off the market. It is a product that must be made safe before it can return. Companies face strong incentives to avoid recalls because they create public relations damage and a competitive disadvantage. Recall shifts incentives. It pushes companies to test carefully, proceed deliberately, and prioritize safety over speed. But those incentives depend on regulation to exist. Without a uniform standard, the market pulls in the opposite direction.

Reframing bans as recalls

The market will not solve this problem on its own. Under current conditions, investing in safer, less addictive design means generating less engagement and losing users to competitors who optimize for engagement instead. No single company can unilaterally raise standards without putting itself at a competitive disadvantage. This is not a failure of individual companies. It is a structural problem that only regulation can solve. Under current conditions, the market rewards the race to the bottom. A public health framework changes that dynamic. It sets a floor that applies equally to every competitor, turning that race around. This is precisely the logic that justified uniform safety standards in automobiles and pharmaceuticals. Companies did not voluntarily make cars safer until they were required to.[78]

This policy brief has two goals. The first, showing what a comprehensive public health regulatory regime for AI companions could look like, has been addressed in the preceding sections. The second, and more urgent, goal is the focus here: existing legislative proposals framed as bans are better understood as recalls, and therefore already fit within that framework.

Recall maps directly onto all three layers of harm. A product subject to recall can be required to add guardrails against self-harm, suicide, and sexual exploitation, addressing the first layer. It can be required to eliminate engagement-maximizing features such as anthropomorphism and sycophancy that foster emotional dependence, addressing the second layer. And it can be required to demonstrate that it does not disrupt social development in children or erode social skills in adults before it returns to market, addressing the third layer.

76 Adam Raviv, "Over-the-Air Recalls Carry Legal Risks for Auto Manufacturers," Bloomberg Law, June 9, 2025, https://news.bloomberglaw.com/us-law-week/over-the-air-recalls-carry-legal-risks-for-auto-manufacturers.

77 New FDA Initiative Reveals Common Reasons for Drug Application Rejection." Applied Clinical Trials, July 11, 2025. https://www.appliedclinicaltrialsonline.com/view/new-fda-initiative-reveals-common-reasons-drug-application-rejection;; Study Shows Impact of Complete Response Letters on Approval Times, (Regulatory Affairs Professionals Society, 2013), ; Bridget Silverman, Better Luck Next Year: US FDA CRLs May be Rising, (Citeline, 2025) ("By the end of 2024, the FDA approved 11 of the 21 novel agents that received CRLs in 2023").

78 See generally John D. Graham, Auto Safety: Assessing America's Performance (Westport, CT: Praeger, 1989).

Some measures that function as recalls already exist. Current proposals that restrict minors' access to AI companion bots function as product recalls by removing unsafe AI bots from the market for minors. They do not prohibit bots categorically. For example, they do not prohibit a customer service bot that provides information about a phone plan. Neither do they prohibit travel services using bots to create itineraries and provide travel information. They do prohibit bots that are unsafely designed. Some of these measures generally prohibit anthropomorphized AI companions that simulate companionship.[79] Others similarly prohibit AI companions simulating companionship but go further, listing specific unsafe features designed to create interpersonal attachment, such as bots that use emotional appeals or retain personal user information in memory.[80]

The result is not that a product is permanently barred. It is that the product cannot return to market unless it is redesigned to meet safety standards. While currently measures operating as recalls apply only to children, recall authority can also extend to products used by adults. It may not be as comprehensive. Under such authority, bots that operate without guardrails, encourage suicide, or induce psychosis would not be merely prohibited with their operators subject to sanctions; they would be taken off the market.

A common objection to national regulation of AI companion bots is that companies will simply operate from other countries, making domestic restrictions ineffective. This concern is real, but overstated. The most practical enforcement mechanism focuses on distribution. Apple and Google distribute the vast majority of apps to American users through their app stores. Requiring those platforms to remove non-compliant AI companion apps from their U.S. storefronts would create a significant barrier to access regardless of where a company is headquartered. This approach is already used in other regulatory contexts. The EU's General Data Protection Regulation applies to any company serving European users, and app store enforcement has become an important compliance mechanism.[81] Payment processing restrictions offer a parallel tool. The United States has used payment processor pressure to limit access to online gambling platforms.[82] Neither mechanism is flawless. But the goal of regulation is not to achieve perfect compliance. It is to change the incentive structure for the vast majority of companies seeking access to the American market.

A separate objection holds that age restrictions are unenforceable because minors will simply evade them. While this concern is legitimate, regulation demands that companies design their products so that exclusion, rather than access, is the default for minors. Age verification is a mandatory design requirement under this framework, and a platform that fails to implement it is operating out of compliance regardless of whether any individual minor succeeds in gaining access. More importantly, a product that reaches minors without adequate safeguards would be subject to recall. The goal of age-based regulation is not to make evasion impossible but to ensure that noncompliance with design requirements carries the same consequences as noncompliance in any other product safety context.

Regulation of AI companion bots through a public health framework can withstand First Amendment and Section 230 scrutiny. Courts addressing this question so far have declined to treat AI companion output as protected speech.[83] Further, in the first social media addiction trial, the jury found Meta and YouTube liable based on defective design rather than speech, confirming that courts are prepared to treat harmful platform architecture as a functional product rather than expression.[84] The design features targeted by

79 GUARD Act, S. 3062, §§ 3(1), 6.

80 New York State Senate, Senate Bill S9051, 2025–2026 Legislative Session, "Prohibits artificial intelligence chatbots from using features which are considered unsafe for minors," introduced January 27, 2026, https://www.nysenate.gov/legislation/bills/2025/S9051.

81 See Julia Krämer, "The Death of Privacy Policies: How App Stores Shape GDPR Compliance of Apps," Internet Policy Review 13, no. 2 (April 2, 2024) https://policyreview.info/pdf/policyreview-2024-2-1757.pdf (evaluating the contributions and deficiencies of app stores as compliance mechanisms).

82 Unlawful Internet Gambling Enforcement Act of 2006, 31 U.S.C. §§ 5361–5367.

83 See Garcia v. Character Techs., Inc. No. 6:24-cv-01903. U.S. District Court for the Middle District of Florida. October 22, 2024.

84 K.G.M. v. Meta Platforms, Inc., JCCP No. 5255 (Cal. Super. Ct. L.A. Cnty. Nov. 5, 2025) (order denying defendants' motions for

this policy brief's proposed framework, including features inducing anthropomorphism, sycophancy, and engagement-maximizing architecture, are functional rather than expressive. Neither Section 230 nor the First Amendment categorically immunizes platforms for causing content-agnostic harms through defective product design.[85] Further, even where some speech protection might apply, courts evaluate content-neutral regulations of functional software design under the deferential intermediate scrutiny standard, a standard that public health regulation of AI companions can readily satisfy.[86]

An important advantage of bans framed as recalls is that they buy time. Use of AI companion bots is currently growing exponentially.[87] Prohibitions depend on enforcement and could expose users, especially children, to years of harm before any relief arrives. Further, once children use AI companions regularly, it becomes much harder to change addictive designs. Even adults in intimate relationships with AI companions have reacted with great distress when companies released new versions that altered their bot's personality or behavior.[88] The longer the industry waits to make changes, the bigger the societal damage, and the harder those changes will be for users to absorb. New norms will have formed, and new dependencies will have developed on specific AI companions and on AI companions as a whole. Delay is not a neutral choice. Every month without action is a month in which the entire population of users serves as the test population. Recall allows harmful products to be withdrawn and safety-tested on a limited basis before they return to market.

Analyzing these restrictions through the lens of recall clarifies what they actually accomplish. They restructure incentives, making it essential for technology companies to invest in safety testing before bringing products to market. The possibility of a recall further incentivizes manufacturers to maintain good data, conduct internal research, and mitigate harms as quickly as possible, including by voluntarily recalling the product.[89]

Conclusion

AI companion bots are already producing documented harms. Without guardrails, they are driving users toward suicide and self-harm and sexually exploiting minors. Their addictive design manipulates users into attachment and emotional dependence. They also pose serious risks to children's social development and to the social skills of adults. This policy brief has made two arguments. First, that a public health framework centered on pre-market approval as the baseline and recall as the most urgent tool is the appropriate regulatory response. Second, that existing legislative proposals framed as bans are better understood as recalls and already fit within that framework.

Pre-market approval shifts the burden where it belongs: onto companies to demonstrate safety before products reach users. Recall addresses the products already on the market. Together, they cover the full lifecycle of AI companion bots and address all three layers of harm. This framework extends regulatory scrutiny beyond intent to include impact. AI companions are not designed for medical purposes, but they demonstrably harm public health. Regulation should follow from that impact, not solely from the intentions

summary judgment); Bobby Allyn, "Jury Finds Meta and Google Negligent in Social Media Harms Trial," NPR, March 25, 2026, https://www.npr.org/2026/03/25/nx-s1-5746125/meta-youtube-social-media-trial-verdict.

85 Brett M. Frischmann and Peter Ormerod, "Regulating Manipulative Design is Not Preempted by CDA 230 or the First Amendment," SSRN Working Paper, October 10, 2025, last revised April 2, 2026, https://doi.org/10.2139/ssrn.5587430.

86 See Lawrence, "Public Health Law's Digital Frontier."

87 Grand View Research, "AI Companion Market Size, Share & Trends Analysis Report," accessed April 29, 2026, https://www.grandviewresearch.com/industry-analysis/ai-companion-market-report.

88 Reddit, "r/replika," accessed April 29, 2026, https://www.reddit.com/r/replika/.

89 Wei-Pin Chen et al., "Regulatory Insights From 27 Years of Artificial Intelligence/Machine Learning–Enabled Medical Device Recalls in the United States: Implications for Future Governance," JMIR Medical Informatics 13 (July 2025): e67552, https://doi.org/10.2196/67552;Ravi Patel et al., "A Retrospective Regulatory Analysis of FDA Recalls Carried out by Pharmaceutical Companies from 2012 to 2023," Drug Discovery Today 29, no. 6 (2024): 103993, https://doi.org/10.1016/j.drudis.2024.103993.

of their designers. Neither pre-market approval nor recall represents a final determination. A product denied pre-market approval can be redesigned and resubmitted. A recalled product can return to market once it meets safety standards. The framework creates incentives to build safer products, not permanent barriers to market entry.

A public health framework does more than protect users. It changes the competitive dynamics of the entire industry. Right now, safety is a liability. A company that invests in less addictive, less manipulative design loses engagement and market share to competitors who do not. No company can solve this alone. Uniform regulation sets a floor that applies to every competitor equally. It transforms the incentive structure of the industry, making safety a competitive asset rather than a penalty.

The loneliness crisis is already a public health emergency, and the evidence suggests AI companions are contributing to it. Addictive design fosters emotional dependence, and users who become dependent spend increasing time with bots and less time in real relationships. A second and distinct risk is that children who do not develop social skills and adults whose skills erode over time will become less able to form and sustain the human connections they need. Both pathways lead to loneliness, and loneliness carries well-documented consequences for physical and mental health, including increased risk of premature death, heart disease, stroke, anxiety, and depression. The full consequences of these trajectories may not be visible for years. The public health framework this policy brief has proposed is designed to address them systematically. The time to act is before dependencies deepen and the costs of intervention rise further.

Many harmful bots are already on the market, making the most urgent step action on existing legislative proposals. Those proposals are framed as bans, but that framing mischaracterizes what they actually do. They do not prohibit AI companion bots categorically. They prohibit bots designed to simulate human relationships to interact with children. Understood as recalls, these proposals fit squarely within the public health framework this policy brief proposes. A safer bot can remain on the market. An unsafe one must be redesigned before it can return.

ACKNOWLEDGEMENTS

I would like to first and foremost thank Rebecca Winthrop for her encouragement and support for this project. I am also grateful to thank Brendan Bouffard, Ravi Iyer, and Frank Pasquale for their thoughtful feedback. I would further like to thank the entire Brookings team, particularly Izzy Taylor and Wafa Abdein. I am also grateful to my research assistants: Gia Gencic, Janine Lechien, and Hetvi Shah for their dedicated work.

Support for my work on this policy brief was provided by The Rockefeller Foundation through a Bellagio Center Residency. My views and content of this brief do not reflect the views or position of The Rockefeller Foundation.

I also gratefully acknowledge sabbatical support from the Seton Hall University School of Law.

BROOKINGS

1775 Massachusetts Ave NW,
Washington, DC 20036
(202) 797-6000
www.brookings.edu