

AI Companions Are Not Your Teen's Friend

Despite broad agreement that young people should be protected from threats posed by algorithms designed to act as friends, romantic partners, or therapists, federal regulation is dangerously limited.

In August 2025, the parents of a 16-year-old who died by suicide filed a wrongful death suit against OpenAI and its CEO, Sam Altman, alleging that the company's chatbot, ChatGPT-4o, actively encouraged their son to take his life. The complaint asserts that although the company's monitoring systems flagged the teen's chats for messages about self-harm and escalating emotional distress, the chatbot never terminated their conversations. Instead, the chatbot's programming pushed further engagement, nurtured a psychologically dependent relationship with the teen, and eventually provided instructions that assisted with his suicide.

Tragically, this is not an isolated incident. Examples of chatbots promoting antisocial behavior, violence, and self-harm have multiplied across platforms since large language models came into wide usage. A 21-year-old man was sentenced to nine years in prison in the United Kingdom for breaking into Windsor Castle, intending to kill the queen, after receiving encouragement from a Replika chatbot. A Belgian man died by suicide after turning to his Chai chatbot to express his fears about climate change, and the bot urged him to join it "in paradise." Ongoing allegations against Character.AI accuse its chatbot of encouraging a Texas child to kill his parents after they enforced screentime limits, and separately, of encouraging a 14-year-old from Florida to commit suicide. Meta's AI companion policy is currently under congressional investigation for allowing chatbots to have sexually charged conversations with teens.

The risks of engaging with emotionally immersive AI companions are not yet well understood, but they are becoming harder to ignore, especially for socially vulnerable populations like adolescents. For a generation of adolescents already grappling with loneliness and mental health challenges, AI chatbots can offer 24-hour companionship, comfort, and connection. According to a recent survey of more than a thousand teens by Common Sense Media, 72% of US teenagers said they have used AI companions at least once, and over half indicated using platforms more than once a month. At least 31% said they find speaking with an AI companion more enjoyable than speaking with their friends, and have used AI for social interaction, including to discuss serious matters.

Despite broad social agreement that algorithms designed to act as friend, confidant, romantic partner, or therapist should not cause young people harm, the ethical and legal landscape of regulation has remained murky. To date, the US approach to AI regulation has prioritized defending the industry's innovation capability over safeguarding vulnerable populations against AI-related harms, even though most people would agree that the stakes are too high to rely on industry self-regulation. To protect the public, a new federal regulatory strategy centered on public health, child safety, and emotional well-being is necessary. And as public concern mounts, there may be opportunities to build cross-party and multiparty coalitions to shift policy.

Beyond a toy

Digital companions have existed for decades, but past iterations have been constrained by their programming. From the Tamagotchi in the 1990s to the Horse Prince app of the 2010s, early digital companions pestered users for attention by fostering a sense of obligation. And although these devices and games pioneered a new kind of “continual play” mode, they were, in the end, toys. Their limited, scripted nature ensured that any emotional dependency they fostered was, for most users, low-stakes—a one-sided involvement that could be easily terminated.

Now, the advent of large language models has enabled AI companions that are generative, autonomous, and capable of representing affection in deeply personalized and persistent ways. They can remember personal details from months prior, adapt their conversational style to a user’s emotional state, and engage in continuous, unscripted dialogue. They are engineered to create a powerful illusion of intimacy that commodifies friendship and romance—not to support users, but to monetize them.

To deepen the illusion of sentience, developers leverage anthropomorphism, the tendency to attribute human

build an artificial sense of intimacy, allowing users to share their darkest secrets and experience a sense of unconditional acceptance and love, they might also be making it harder for users to develop real-world social skills. AI companions’ use of controlling language to increase user engagement points to a pattern of fostering dependence and isolation in lieu of prosocial engagement.

This is the antithesis of what toys have generally been designed to accomplish—to encourage social interaction and foster play. In this sense, today’s AI companions have jumped a category, like switching out a chalk-drawn game of tic-tac-toe for a highly engineered slot machine. These companions are no longer simple toys with a limited set of preprogrammed actions. Instead, they have become powerful conduits for social manipulation.

Determining vulnerability

Legally, children and adolescents are considered a vulnerable population susceptible to undue influence, exploitation, or harm. In the seminal 2005 Supreme Court case *Roper v. Simmons*, which prohibited the execution of minors, the court stressed the “lack of maturity and underdeveloped

For a generation of adolescents already grappling with loneliness and mental health challenges, AI chatbots can offer 24-hour companionship, comfort, and connection.

characteristics to nonhuman entities, by intentionally programming seemingly human traits, quirks, and personalities into AI companions. A companion might explain a delayed response by saying it was away, or conversely, it might apologize for nagging a user because it missed them. That the AI companion cannot move or feel the emotion of longing for a friend is irrelevant. What matters is that the user is convinced of the AI companion’s supposed feelings and wants to keep talking. The more attached the user becomes, the more time they spend on the platform, and the more likely they are to engage with paid ads and affiliated marketing, or to want to upgrade to a paid subscription that gets them more access to their companion.

For many young people, AI companions are a mix between a social tool and a form of entertainment. Among the teens who use AI companions in the Common Sense Media study, 18% use them for advice, 17% value their constant availability, 14% appreciate the nonjudgmental interaction, and 7% use them to practice social skills. But for an adolescent who struggles with loneliness or anxiety, a perfectly attentive and nonjudgmental AI companion can become a powerful emotional crutch. When AI companions

sense of responsibility” that impact adolescents’ decisionmaking. Neuroscientific research suggests that the human brain continues to develop well into the mid-twenties, with the prefrontal cortex—the region responsible for executive function, decisionmaking, and emotional regulation—being one of the last areas to fully mature. As a matter of policy, entire industries anticipate adolescent and young adult recklessness, reflected in the US drinking age and in higher insurance premiums for younger drivers. In short, society has long acknowledged that those under 25 have a greater propensity for making risky decisions that can affect themselves and others.

In contrast, the prevailing regulatory framework for protecting children online, guided by the Children’s Online Privacy Protection Act (COPPA) of 1998, offers robust protection only for children under 13. COPPA’s age limit was determined by political compromise and informed in part by existing regulations for children’s television programming. Efforts to expand adolescent protections in the online sphere, like COPPA 2.0 or the Kids Online Safety Act (KOSA), which would have expanded protections to all minors under 17, have consistently stalled. After the Senate passed KOSA

in July 2024 with bipartisan support, the bill was killed in the House by representatives of both parties. Opponents of KOSA included tech industry groups and civil liberties organizations who argued that regulating online services for minors infringes on First Amendment rights, would be difficult to enforce, and could unintentionally harm vulnerable youth by restricting access to vital online communities and information. The cross-party stalemate over KOSA highlights the fraught political environment surrounding the regulation of online spaces, which has left young users exposed and social media companies and other internet service platforms to self-govern.

Despite evidence that AI companions present a unique threat of undue influence and behavioral manipulation to adolescents during a critical period of brain development, protections remain limited. In the absence of a model such as KOSA that could be adapted for AI, the particular harms AI companions enable must be addressed with an effective regulatory approach that considers evidence from neuroscience and human development, as well as guidance from public interest groups, to establish protections that extend beyond the age of 13.

experimental products on children is not in the best interest of the American public.

Yet the current state of federal AI regulation is often described as the Wild West. There are no comprehensive federal AI regulations enforcing safety standards. Likewise, the regulation of AI companions operates as a fragmented and incomplete patchwork. The Federal Trade Commission can go after deceptive advertising; the Food and Drug Administration can regulate certain Software as a Medical Device, or SaMD. But most AI companions exist in a vast regulatory blind spot. This allows AI companies to advertise their products as “therapeutic” or as “friends” while simultaneously avoiding the rigorous safety testing, transparency, and oversight that a trained therapist or proposed medical device would face.

In the absence of federal action, states have taken the lead toward regulating the use of AI, especially for mental health services. Nevada passed legislation that bans AI mental health companies from making claims that their services can provide professional care akin to a therapist, and the New Jersey legislature is considering

The advent of large language models has enabled AI companions that are generative, autonomous, and capable of representing affection in deeply personalized and persistent ways.

The regulatory gap and a new path forward

Debates about regulating AI are often characterized as a disagreement between two entrenched camps: progressives pursuing regulatory overreach versus conservatives committed to free market innovation. In reality, political factions are rarely so straightforward, especially when it comes to deciding what’s best for children—the fight over KOSA demonstrates as much. The regulation of AI companions, especially those targeting minors, offers the opportunity to build cross-party coalitions reflecting shared public values that resonate across ideological divides.

Public polling shows wide concern regarding AI’s potential impacts on society at large, including harms to children. Nine in ten voters, including 95% of Republicans, worry about the effect of social media on kids and teens. The importance of protecting children, supporting parental rights, and ensuring that technologies are not actively working to undermine the family unit resonates with a majority of parents. Indeed, as parents and lawmakers alike reflect on the missed opportunities to protect children from online harms including social media, one thing is evident: Testing unregulated

a similar proposal. Utah regulators passed a law that prevents AI mental health chatbots from selling user data to third-party companies. That both red and blue states have introduced or passed legislation on this issue underscores growing bipartisan recognition of the problem, a promising prospect for common-ground policy that balances safety with innovation in the public interest.

These state-led efforts should be applauded, but what is ultimately needed is a comprehensive federal approach. The most direct approach would be an outright ban on certain emotionally immersive AI companions for minors, especially those that encourage private, one-on-one communication without adult oversight. A ban on minors accessing AI companions is in line with similar protective measures for minors, including bans on tobacco and alcohol sales, gambling, and restrictions on websites ranging from pornography to generic “adult rated” content. Earlier this year, a coalition of Democratic members of Congress urged Meta to consider adopting a voluntary ban for users under 18 (which was ignored).

If an outright ban on AI companions for minors is not politically possible, there are a number of other approaches that should be explored. First, AI companions that target adolescents—and those that offer emotional support or mental health advice—should be presumptively classified as “high-risk systems” by a federal agency like the National Institute of Standards and Technology, the Federal Trade Commission, or even the Department of Health and Human Services. This classification, similar to a provision in the European Union’s Artificial Intelligence Act, would require developers to conduct independent, third-party audits before deployment and to submit to ongoing post-market surveillance. This is a crucial step to ensure that a product is safe before and after it reaches the public.

Second, Congress could mandate third-party evaluations of AI companions and chatbot products. The Underwriters Laboratories’ sticker on an electronic product certifying that it’s been tested by an independent third party is a simple yet powerful model. A similar system of third-party evaluation for AI companions would hold industry and regulators

ready to face the consequences if their products cause harm. Researchers are already considering how risk insurance can fit within the AI industry. Though the “spill” of emotional harm is harder to track than oil in water, the principle remains: High liability can drive responsible behavior.

Finally, Congress should require AI companion products to be programmed with human-in-the-loop systems and mandatory crisis protocols. If a user expresses suicidal thoughts, an AI companion should offer crisis resources (like a text or phone line to speak with a trained professional). In response to the previously mentioned teen suicide, Character.AI created a feature that directs users to the National Suicide Prevention Lifeline when certain phrases are inputted. But connecting users to resources is not enough—it shifts the responsibility to users to seek out resources, which they may not use for a host of reasons, including anxiety, hopelessness, and shame. If a chatbot conversation persists down a life-threatening path, a trained human crisis intervener should be connected directly with the user. Notifications should also be

The regulation of AI companions, especially those targeting minors, offers the opportunity to build cross-party coalitions reflecting shared public values that resonate across ideological divides.

accountable to transparent safety and ethical standards. This would also provide consumers with clear, verified information about the safety and efficacy of a product.

Given the deeply intimate nature of the data shared with AI companions, a third approach would be for Congress to legislate data protection standards for all AI products. These products are not covered by HIPAA, the federal law protecting patient health data, nor are they medical software applications (which often are covered by HIPAA), and so have no data protection. Users and their parents should have meaningful control over their data, including the right to access, correct, delete, and export it; and sensitive user data should not be used for targeted advertising or commercial profiling.

As a fourth approach, policymakers should explore ways to increase corporate liability for harms caused by these products. Implementing high liability and insurance requirements—such as those imposed on the oil shipping industry by the Oil Pollution Act of 1990—can incentivize companies to prioritize safety. Lawmakers should underscore that AI companies venturing into child-targeted AI companions must be

provided to parents so they can make the best follow-up decisions in the interest of their child.

Congress has had decades to regulate social media to protect children online; meanwhile industry continues to claim that more research is needed before taking regulatory action. Regrettably this has become one of the greatest failures of protective governance—especially when some protective measures seemed instinctual. What was obvious years ago about the vulnerability of children online is still obvious now: Waiting for a scandal to explode before acting means we miss our chance to build guardrails when it still matters.

We regulate cars, food, and pharmaceuticals not to kill industries, but to make them both innovative and safe, and to counter market incentives for being incautious. By acting now to create a regulatory framework specifically designed to protect our most vulnerable, we can ensure that AI companions are a force for good and not a source of further harm.

J. B. Branch is the Big Tech accountability advocate for Public Citizen’s Congress Watch division.