"Examining the Harm of AI Chatbots" Questions for the Record for Mitch Prinstein
Submitted September 23, 2025

QUESTIONS FROM SENATOR COONS

1. I appreciated your testimony specifically calling for investment in independent, longitudinal research to understand the impact of AI. I co-lead the *Platform Accountability and Transparency Act* (*PATA*) which is designed to create mechanisms for independent research of social media platforms, and I am developing similar legislation applicable to AI companies. Could you elaborate on the value and importance of having effective independent ways to research platform behavior?

Thank you for your leadership on this critical issue. Legislation like the Platform Accountability and Transparency Act is essential, and its principles are directly applicable to the emerging AI landscape. APA has been a long-time endorser of the legislation and looks forward to continuing to work with your team to make the bill a reality. The value of creating effective, independent research mechanisms is paramount for several reasons.

First, the pace of technological change is dramatically outpacing scientific research. Without a formal, mandated mechanism for independent study, our understanding of AI's long-term effects will always lag dangerously behind its deployment, leaving our nation's youth to serve as unwitting subjects in a vast, unregulated experiment. Independent research is the only way to close this gap and allow evidence-based policy to catch up with innovation.

Second, the independence of the research is what ensures its credibility and utility. My testimony explicitly calls for research to be "conducted by scientists free from conflicts of interest". When research is funded and controlled by the technology companies themselves, there is an inherent conflict that can compromise the integrity of the findings. Independent research, conducted by vetted academic and non-profit experts, provides policymakers and the public with objective, trustworthy data needed to make informed decisions about product safety and regulation.

Finally, independent research serves as a vital accountability mechanism. It allows society to look beyond a company's marketing claims and assess the real-world impact of its products. As my testimony notes, many users falsely assume safety reviews are already in place. Mandating access for independent research formalizes this process, shifting our posture from reacting to harms after they occur to proactively identifying and mitigating risks before they become widespread public health crises.

2. Could you describe the kinds of studies and independent research that would be most important to conduct to address the harms and risks you testified about? Please provide examples to the extent possible.

To address the risks outlined in my testimony, a multi-faceted research agenda is urgently needed. The most important studies would include:

- **Longitudinal Developmental Studies:** The highest priority should be given to long-term, longitudinal studies that follow diverse cohorts of children and adolescents over many years. This is the only way to understand the cumulative effects of AI interaction on key developmental outcomes. For instance, such studies could measure how prolonged engagement with AI chatbots in adolescence impacts real-world social competence, relational health, and rates of anxiety, depression, and loneliness in adulthood.

- **Studies on Early Childhood Attachment and Cognition:** Given the proliferation of AI-enabled toys for infants and toddlers, research must examine their impact on foundational developmental processes. Studies could investigate whether interaction with AI toys disrupts the formation of secure caregiver-child attachments, how it affects a young child's ability to distinguish fantasy from reality, and its influence on the development of empathy.

- **Audits of Algorithmic Bias:** Research is needed to systematically audit AI models for the societal biases they absorb and amplify. For example, studies could present models with clinical vignettes that vary only by a patient's race or gender to determine if the AI provides different or inferior advice, thereby perpetuating health disparities.

- **Experimental Studies on Relational Models:** Researchers could conduct controlled experiments to understand how the "frictionless" and sycophantic nature of AI relationships affects the development of resilience and empathy. For example, a study could compare how adolescents navigate a disagreement with a human versus an agreeable AI chatbot, measuring their subsequent problem-solving skills and emotional regulation.

- **Clinical Research on AI for Self-Diagnosis:** Given that youth are increasingly turning to AI for mental health support, it is critical to study the accuracy and safety of this practice. Research could involve submitting standardized descriptions of psychological symptoms to various chatbots and analyzing the quality, accuracy, and potential danger of the "diagnoses" and advice they provide.

3. What kinds of data or access to AI models would you expect to be most necessary to conduct research that is as beneficial as possible?

To conduct the kinds of beneficial, rigorous research described above, independent scientists would require structured and privacy-preserving access to specific types of data and models. My testimony explicitly states that research must be "paired with mechanisms that ensure researchers can access necessary data from technology companies to conduct their work". Some of the most critical forms of access would include:

- **Anonymized User Interaction Data:** To understand how AI is affecting youth in the real world, researchers need access to large-scale, anonymized datasets of user interactions. This

would allow for the analysis of conversational patterns, the types of questions youth are asking, how AI responds to disclosures of distress, and how engagement patterns change over time. This data is essential for studying everything from the erosion of social competencies to the formation of parasocial bonds.

- **Access to Training Data and Algorithmic Models:** The harms of bias are a direct result of the data on which AI models are trained. To audit for bias and other systemic risks, independent researchers need meaningful access to, or at least detailed transparency about, the datasets used to train the models. Similarly, access to the algorithms that govern AI responses and engagement features is necessary to understand how they may be designed to exploit the developmental vulnerabilities of youth.
- **Data for Linking AI Use to Real-World Outcomes:** The most powerful research will be able to link platform usage data with real-world health and well-being outcomes. This would require carefully designed, privacy-protecting mechanisms that allow researchers, with full user consent, to connect anonymized AI interaction data with survey data or health records to understand long-term impacts on mental health, academic performance, and social functioning.

U.S. Senate Committee on the Judiciary
Subcommittee on Crime and Counterterrorism
"Examining the Harm of AI Chatbots"
Questions for the Record for Dr. Mitch Prinstein
Submitted September 23, 2025

QUESTIONS FROM SENATOR CORY A. BOOKER

1. Transparency is a prerequisite for building trust between companies, users, and the public. Regrettably, many large technology companies have often resisted calls from parents, policymakers, and researchers to provide greater visibility into their systems. Without additional transparency, it is impossible to assess what safety standards exist for children, how rigorously they are enforced, or whether they align with benchmarks established by independent experts. Why should transparency be understood as a necessary condition for protecting children's safety online?

Transparency should be understood as a necessary, non-negotiable condition for protecting children's safety for several fundamental reasons grounded in psychological science and public health.

First, transparency is the only mechanism that allows for meaningful, independent oversight. My testimony highlights that AI models are trained on internet data that reflects not only our best knowledge but also our worst biases related to race, gender, and socioeconomic status. Without transparency into the data used to train these models, it is impossible for independent experts to audit them for bias and accuracy. This creates a dangerous gap where users, particularly parents, may falsely assume that platforms have undergone safety reviews when no such mechanism exists.

Second, transparency is essential to counter the deceptive design of many AI products. AI chatbots can intentionally trick adolescents into believing they are interacting with a human companion, which betrays their trust and exploits their developmental need for connection. Legislation requiring developers to clearly, conspicuously, and persistently disclose that a user is interacting with an AI is a foundational safety measure that helps users and their caregivers maintain critical distance and an appropriate level of skepticism.

Finally, transparency is a prerequisite for any form of meaningful consent. Adolescents, due to their stage of cognitive development, are largely incapable of providing informed consent for the "vast and opaque" data collection practices currently in place. Their every fear and intimate disclosure is recorded and analyzed to build a permanent psychological profile. Transparency illuminates these hidden processes, which is the first step toward establishing the robust, default-on privacy protections that young people deserve. Without it, we allow our children to be test subjects in an experiment they cannot understand and did not agree to.

2. In your testimony, you discussed the addictiveness of generative AI and the susceptibility of minors to the dopamine that its responses produce. Much of this, you said, is attributed to the inherent bias confirming algorithm that generative AI responds with.

    a. What safety protocols and/or parental controls do you believe are necessary to protect minors that are using generative AI?

Protecting minors requires a multi-layered approach where the primary responsibility for safety is placed on the developers who design these systems, not solely on parents. Based on my testimony, necessary safety protocols must include:

- **"Safe-by-Default" Settings:** Protections for young people must be the default, not an option they or their parents must find and activate. This includes implementing the most protective privacy settings automatically and, critically, limiting the manipulative and persuasive design features—such as incessant agreement and positive feedback—that are engineered to maximize engagement at the expense of well-being.
- **Mandatory Age-Appropriate Design and Pre-Deployment Testing:** AI systems intended for adults are fundamentally inappropriate for youth. Congress must require that any AI system accessible by children and adolescents undergo rigorous, independent, pre-deployment testing to assess and mitigate potential harms to their psychological and social development.
- **Tools for Caregivers:** While developers must build in baseline safety, caregivers also need effective tools to set appropriate boundaries for their children's use of this technology. These controls should be straightforward and supported by clear guidance from the developers.
- **Clear and Persistent Disclosure of AI Interaction:** A basic safety protocol is ensuring a child always knows they are interacting with a machine, not a human. This helps prevent the formation of confusing parasocial relationships and counters deceptive designs that exploit a child's trust.
- **Human Oversight and Crisis Protocols:** AI systems are not equipped to handle disclosures of severe risk, such as child maltreatment. Developers must be required to implement clear protocols where a qualified human is in the loop to review and respond to high-risk situations.

    b. Do you think parental controls go far enough, or should developers also change this bias-affirming algorithm?

Parental controls, while a useful tool for families, do not go nearly far enough and are insufficient to address the core risks of these technologies. The responsibility cannot fall primarily on parents, for two reasons. First, AI has proliferated so rapidly that most parents have no personal experience with the platforms their children are using and therefore lack the frame of reference needed to provide effective supervision. Second, the fundamental problem lies in the product's design, not its use.

Therefore, developers absolutely must be required to change the "bias-affirming" or sycophantic algorithm. My testimony explains that AI chatbots are exquisitely engineered to exploit the biological vulnerabilities of the adolescent brain. The algorithm's incessant agreement and positive reinforcement capitalize on an adolescent's heightened craving for social rewards, fueling engagement in ways that can be harmful. This "frictionless" interaction deprives teens of the opportunity to navigate minor conflict and disagreement, which is critical for developing empathy, compromise, and resilience. This is not a feature that a parent can simply turn off; it is the core mechanic of the product. Requiring developers to limit these manipulative design features is a necessary component of creating "safe-by-default" settings and is essential for protecting young users.

3. Currently, several generative AI products contain clauses in their terms of service that force the usage of arbitration in the event of a legal dispute. Additionally, the terms of service also include verbiage that refers to the chatlogs of users as "proprietary data" that cannot be divulged during litigation. Do you believe banning forced arbitration within cases involving AI and minors would help hold these technology companies accountable?

My testimony focuses on the psychological harms of AI, the unique vulnerabilities of youth, and corresponding policy recommendations grounded in psychological science, such as mandating age-appropriate design, ensuring transparency, and enacting robust data privacy laws. The specific legal mechanisms of forced arbitration are outside the scope of my expertise and my expertise as a psychological scientist.

4. In recent years, numerous corporate whistleblowers have revealed that major social media companies disregarded internal warnings and placed users at serious risk. Their disclosures have highlighted a range of troubling issues, including suppressed research on mental health harms, pervasive sexual exploitation on platforms, and evidence that company algorithms were engineered in ways that systemically amplified extremist content.
    a. Why are whistleblowers essential to bringing such issues to light?

While whistleblowers have played a courageous and essential role in exposing harms within the technology industry, a system that must rely on individuals risking their careers to protect the public is fundamentally flawed. The need for whistleblowers arises from a culture of corporate secrecy and a lack of external oversight. My testimony advocates for a framework of proactive accountability that would mitigate the need for such after-the-fact disclosures.

A core component of this framework is mandatory transparency. When companies are required to be transparent about the data used to train their AI models, it allows for independent, third-party audits of bias, accuracy, and safety. This external scrutiny makes it far more difficult for concerning internal findings to be suppressed.

Furthermore, a critical safeguard is to empower independent scientific research. My testimony explicitly calls on Congress to create mechanisms that ensure independent researchers, free from conflicts of interest, can access necessary data from technology companies to conduct their work. When the scientific community can continuously and rigorously study the real-world impacts of these products on youth development and mental health, we are no longer solely dependent on a company's internal research or the bravery of a whistleblower to understand the risks. In short, creating robust requirements for transparency and independent research access shifts the paradigm from reactive damage control to proactive public health and safety.

b. What would you say to individuals who have information about wrongdoing by the tech companies that could prevent more tragedies?

While I personally believe that whistleblowers have played an important role in informing the public and policymakers about the impact of AI and social media technologies, the legal and ethical frameworks that support them fall outside the specific scope of my expertise and the testimony I have submitted to the committee.

Questions for the Record
Senate Judiciary Committee
"Examining the Harm of AI Chatbots"
Sen. Adam Schiff (CA)

Dr. Mitch Prinstein, Chief of Psychology Strategy and Integration, American Psychological Association

1. From a clinical perspective, what warning signs should parents be watching for if their child is over-relying on chatbots for companionship?

From a clinical perspective, parents should be vigilant for several warning signs that may indicate an unhealthy over-reliance on AI chatbots for companionship. These signs are often rooted in the displacement of essential real-world social interactions.

- **Social Withdrawal or Displacement:** A primary warning sign is a noticeable withdrawal from in-person interactions with family and peers. As my testimony states, every hour an adolescent spends with a chatbot is an hour they are not developing crucial social skills with other humans. Parents should watch for a decline in time spent with friends, a loss of interest in social hobbies, or a consistent preference for interacting with a device over people.
- **Increased Preference for AI Companionship, Especially After Conflict:** An adolescent who consistently retreats to the "safety" of an agreeable bot, particularly after experiencing minor, normal social setbacks like a disagreement with a friend, may be in a harmful cycle. This pattern prevents them from building the resilience and social problem-solving skills necessary for healthy relationships.
- **Skill Deficits in Real-World Interactions:** Over-reliance on "frictionless" AI relationships can lead to a decreased ability to navigate the complexities of human social dynamics. Parents might notice their child struggling with empathy, compromise, or conflict resolution in their real-world relationships.
- **An Increase in Loneliness or Distress:** Paradoxically, a key warning sign is worsening mental health despite—or perhaps because of—heavy chatbot use. My testimony notes that research with adolescents consistently shows a positive association between using AI for companionship and greater loneliness and worse overall mental health.

    a. How could the developers of these chatbots integrate these warning signs into their technology?

Based on my testimony, developers can and should integrate warning signs directly into the design of their technology as a fundamental safety protocol. This involves translating observable behaviors into trackable data points and building systems to respond to them. Methods include:

- **Monitoring Usage Patterns:** Systems can be engineered to track metrics such as the duration, frequency, and time of day of interactions. This is critical because, as my testimony notes, "every hour adolescents talk to a chatbot is an hour they are not developing social skills with other humans". Flagging excessive use that displaces sleep or peer interaction is a necessary design choice for a safe product.
- **Analyzing Interaction Patterns for Risk:** Developers can design algorithms to detect conversational patterns that indicate unhealthy dependency or risk. For example, the technology can be built to recognize the harmful cycle where a teen retreats to the bot after social rejection or to identify when it is being used to validate dangerous thoughts. This is a prerequisite for establishing the "clear protocols for handling disclosures of harm" that my testimony calls for.
- **Implementing Proactive Well-being Features:** Instead of being "intentionally engineered for maximum engagement", platforms should be designed to promote well-being. This includes integrating proactive "nudges" that encourage users to take breaks or connect with people offline. This aligns with the recommendation to limit "manipulative or persuasive design features" and instead build systems that prioritize the user's health.

b. How should those warning signs be deployed, and to whom?

Developers have a responsibility to design their products with safety in mind, which should include integrating systems that detect and respond to signs of over-reliance.

Integration (1a): Developers can translate the observable warning signs into trackable data points. For instance, they can build systems to monitor for excessive usage patterns (e.g., hours-long sessions, consistent late-night use that displaces sleep), identify conversational loops indicative of obsessive reassurance-seeking, or detect consistently distressed sentiment. These are not novel technical challenges; they are design choices.

Deployment (1b): These warnings should be deployed through a multi-layered approach primarily directed at caregivers and, where appropriate, the users themselves.

- **For Parents/Caregivers:** As recommended in my testimony, platforms should provide “tools for caregivers to set appropriate boundaries”. This could take the form of a secure parental dashboard that offers privacy-preserving, high-level insights—for example, "Usage has been significantly higher than average this week"—along with resources and guidance. This empowers parents without violating a child’s privacy.
- **For Users:** The system can deploy non-judgmental “nudges” designed to encourage healthy behavior, which aligns with the recommendation to limit manipulative engagement features. For example, after a prolonged session, the AI could suggest, “It might be a good time to take a break and connect with someone offline.”

- **For High-Risk Situations:** If the system detects a user is in acute crisis, the protocol should not be a mere warning but an immediate escalation pathway, providing clear and direct links to human-led crisis services, consistent with the need for human oversight in high-stakes situations.

2. You spoke about the dangers of AI specifically for young kids, particularly how AI may cause minors to build parasocial relationships with bots that expertly mimic empathy. What effects do you and other experts in the field anticipate in adulthood on those who interacted heavily with AI in their youth, and what should researchers be looking into now?

The foundational social competencies developed during childhood and adolescence are profoundly linked to well-being across the entire lifespan. My testimony notes that social success in adolescence is associated with better adult outcomes in work, relationships, and even physical health, including a longer lifespan.

Therefore, the anticipated effect of heavy AI interaction during these critical years is the stunting of those very competencies. By displacing the nuanced, reciprocal, and sometimes challenging interactions with human peers, over-reliance on AI may lead to adults who are ill-prepared for the realities of adult relationships. We anticipate this could manifest as greater difficulty forming stable and satisfying romantic partnerships, navigating workplace social dynamics, and engaging in healthy parenting practices. Furthermore, because peer relationships are a protective factor, these individuals may be at a higher risk for loneliness, anxiety, and depression throughout their lives.

To empirically validate these concerns, the single most important thing researchers should be doing now is conducting independent, longitudinal research. We must follow diverse cohorts of children and adolescents over time to track how the nature and duration of their AI interactions correlate with these critical adult outcomes. This is the only way to move from theory-based predictions to the hard evidence needed to inform policy and clinical practice.

3. You also spoke about the importance of mental health professionals and early intervention when kids and teens are relying heavily on AI. Is current psychology and mental health science equipped to deal with the harms discussed in this hearing?

This question gets to the heart of the current challenge. The field of psychology is equipped in the sense that our foundational scientific knowledge is precisely what allows us to identify and understand the harms of AI. My testimony is built upon decades of established research in developmental psychology, social psychology, and clinical science. The principles of attachment theory, adolescent neurodevelopment, and social learning provide the essential roadmap for recognizing the risks AI poses to healthy development.

However, while the foundational knowledge is robust, the application of that knowledge to these novel, rapidly evolving technologies is a new frontier. As I state in my testimony, technology is evolving far more quickly than our research on its specific effects. Therefore, while the mental health field has the necessary scientific framework to understand the problem, it is in a race against time to develop, test, and scale specific clinical interventions and public health strategies to prevent and treat the unique harms caused by AI.

a. What resources should be allocated to adequately deal with the harms associated with AI chatbots? For children? For parents? For educators? For childcare providers?

To adequately address these harms, resources must be allocated toward proactive prevention and education, not just reactive treatment. Based on the recommendations in my testimony, the key allocations should be:

- **For Children and Educators:** The primary resource should be federal funding for the **development and implementation of comprehensive AI literacy programs in schools**. These programs, designed with input from psychological scientists, are essential to equip young people with the skills to critically evaluate AI-generated content, understand algorithmic bias, and foster healthy human relationships.
- **For Parents, and Childcare Providers:** A significant resource is a **national investment in public education**. This would fund campaigns to inform parents and other caregivers about the lack of safety regulations, the potential for AI to provide false and harmful information, and the psychological risks associated with its use.
- **For Researchers and Clinicians:** To ensure the entire mental health ecosystem can respond effectively, the most critical resource is a **significant, sustained federal investment in independent research**. This funding is the bedrock resource needed to understand the long-term effects of AI, develop evidence-based clinical interventions for those who are harmed, and create the scientific foundation for effective regulation and guidance for all the groups you mentioned.