ITIF | INFORMATION TECHNOLOGY & INNOVATION FOUNDATION

How Policymakers Should (and Shouldn't) Address Chatbot Safety for Children

What to Do

✓ Utilize a child flag system

Congress should pass legislation requiring device operating systems to create an opt-in "trustworthy child flag" for user accounts.

✓ Encourage meaningful parental controls to mitigate risks

Platforms should implement and encourage the use of meaningful controls that are easy for parents to understand and access.

✓ Encourage digital literacy

Digital literacy efforts should give children tools to apply their developing critical thinking skills in the context of chatbots.

✓ Require paid content disclosures

Policymakers should require disclosures of when chatbot outputs contain paid sponsored content similar to required disclosures on web searches and social media paid posts.

✓ Create clear and distinct definitions, including carveouts

Policymakers should avoid subjective language that can allow for overregulation and create appropriate distinctions.

✓ Set clear standards for licensed AI professional services

Policymakers should set standards commensurate with those that a human must meet to practice in a certain field.

✓ Address child suicidal ideation conversations

Policymakers and developers should combine crisis-referral protocols with aforementioned standards for licensed AI mental-health services.

✓ Discourage misuse of fictional characters popular with children

Policymakers should ensure that copyright owners can enforce their intellectual property rights to prevent the misuse of their copyrighted characters.

✓ Fund more research

Congress should prioritize funding studies on how AI companions affect different groups, particularly children.

✓ Encourage industry best practices for self-regulation

Policymakers should encourage the AI industry to develop safety standards analogous to those used in other consumer product categories.

What Not to Do

× Require age checks for all users

Alternative methods for age-verification are emerging as more privacy-preserving options, however, defining which chatbot services require verification remains challenging and can lead to over-regulation while taking decision-making power away from parents.

× Restrict content

Content-related requirements either give governments the power to determine what content is harmful to children—a likely First Amendment violation—or force the businesses that offer chatbot services to make those decisions, potentially opening themselves up to legal consequences.

× Perpetuate the "addiction" narrative

"Addiction" is not a scientifically or medically accurate term for chatbot use: Current research does not support that narrative, and policymakers should therefore avoid this trope.

× Rely on warning labels to educate users on harms

Mandating warning labels is a flawed approach that shows concern while not actually providing any meaningful protection. Most users ignore warning labels, and children don't always understand disclosures.

× Ban targeted advertising

Taking targeted advertising revenue away would lead to a greater lack of resources, especially free or low-cost resources which would be especially detrimental to lower-income households.

× Ban chatbots for all minors

Banning chatbots altogether would cut youth off from 24/7 support that safe chatbots and AI companions can provide. Bans would also require these services to implement age-verification mechanisms for all users, which presents privacy risks for adults.

× Scapegoat sycophancy

Not every AI response on a platform will be sycophantic, even on platforms that allow such responses. Platforms could also give users control over sycophantic language through parental settings, letting them decide for themselves how they want their children to engage.

× Regulate time limits

Chatbots should have the same choice as streaming services do as to whether to enable reminders and how often they should occur. Mandating reminders at certain intervals for only certain types of platforms is inconsistent and ineffective government policy.