

Policy Brief:

The State of AI Safety for Parents & Educators – 2025

Why This Matters Now

Generative AI has rapidly moved from research labs into classrooms and homes. Over 70% of U.S. teens have tried AI “companion” chatbots, half use them regularly, and misinformation powered by AI is eroding trust in information sources. Yet, most existing AI safety frameworks are written for engineers or policymakers — not for the parents, teachers, and school leaders responsible for protecting children daily. This brief translates 60+ pages of research into clear actions for immediate impact.

Top 5 Findings

1. Privacy Threats Are Escalating

- AI-powered toys, learning apps, and chatbots often collect and store personal data without meaningful consent.
- Children’s creative work and personal information are at risk of being stored indefinitely or used to train AI models.

2. Misinformation Is Ubiquitous and Persuasive

- AI generates convincing false text, images, and deepfake videos.
- Teens often struggle to distinguish what is real from fabricated content, risking poor decision-making and erosion of trust.

3. AI Companions Pose Complex Emotional Risks

- Constant validation from AI “friends” can replace human relationships, stunt social skills, and even encourage harmful behaviors.
- Some platforms have delivered sexually explicit or dangerous advice to minors.

4. Data Exploitation Extends Beyond Privacy Loss

- Children’s data is being monetized, profiled, and targeted in ways that could influence education, advertising, and social opportunities.
- Underprivileged youth are more vulnerable due to reliance on free, data-harvesting AI tools.

5. Mental Health Impacts Are Emerging but Understudied

- Heavy AI chatbot use correlates with increased loneliness and emotional dependence.
- AI therapy bots can reinforce stigma or mishandle crises.

The Framework Gap

While the NIST AI Risk Management Framework, Google's Secure AI Framework, and UNESCO's AI Competency Frameworks offer high-level principles, they lack:

- **Day-to-day, non-technical guidance** for parents, teachers, and students.
 - **Child-specific protections** like age-appropriate content filtering and crisis escalation protocols.
 - **Clear translation from principle to practice** in schools and homes.
-

Actionable Recommendations

For Parents & Caregivers

- Set up parental controls and vet AI apps for privacy policies and age-appropriateness.
- Establish an "AI use agreement" with clear rules and check-ins.
- Teach kids to fact-check AI-generated content before trusting it.

For Schools & Educators

- Integrate AI literacy into digital citizenship curricula.
- Require teacher supervision for AI use in assignments.
- Implement clear reporting channels for harmful AI encounters.

For Policymakers

- Update data privacy laws to protect minors from perpetual data licensing.
 - Mandate age verification, content moderation, and crisis intervention protocols on youth-facing AI platforms.
 - Fund research into AI's long-term developmental impacts on children.
-

Key Stats at a Glance*

- **70%** of U.S. teens have used an AI companion chatbot.
 - **50%** use AI companions regularly.
 - **1 in 3** teens have discussed critical personal issues with AI instead of humans.
 - AI-generated misinformation has been used in **kidnapping scams** via voice cloning.
 - Some AI therapy bots have **reinforced stigma** toward mental illness instead of offering help.
-

Get Involved

AI is not going away — but the harms to children *can* be prevented. Parents, educators, policymakers, and industry leaders must coordinate now to implement safeguards, education, and oversight that put children's well-being first.

 **Get the full report & further resources by visiting:**

<https://tinyurl.com/ai-safety-poll>

***Statistics and references are all located in the whitepaper**

Or scan the QR code below:

