



The AI Safety Toolkit for Parents

Everything you need to protect your
child from unsafe AI apps in one place.

Introduction

AI is everywhere.

From ChatGPT to Snapchat's My AI, Character.ai, and even Instagram's Meta AI, chances are your child has already encountered an artificial intelligence chatbot.

Some of these tools can be fun and helpful — sparking creativity, supporting homework, or providing quick answers. But we've also seen kids use AI in ways that are unsafe, unhealthy, or emotionally risky.

Recent lawsuits and investigations show that tech companies are only beginning to acknowledge the risks.

Parents can't wait for platforms to catch up. Kids need protection now.

That's where this toolkit comes in.

Inside, you'll find:

- A cheat sheet of the most popular AI apps and what risks they pose
- A simple setup checklist to protect your child in under 15 minutes
- Conversation starters to help you talk with your tween or teen
- A quick quiz to evaluate your child's AI safety
- The best tool parents can use to stay informed without spying



This guide is here to help you feel prepared, not overwhelmed.

AI App Cheat Sheet

Here's how today's biggest AI apps stack up on parental controls and safety.

App	Age Checks	Parental Controls	Self-Harm Alerts	Usage Limits	Content Filters
ChatGPT	Weak (easy to bypass)	Parents can link accounts with teens	Yes	Yes	Limited
Character.AI	Easy to fake age	Weekly activity reports only	No	Yes (soft reminders)	Can be bypassed
Snapchat My AI	Weak (13+)	Can disable via Family Center	Yes	No	Some, but limited
Meta AI (Instagram, WhatsApp)	Weak (easy to bypass)	None	Yes	Hourly reminders only	Some, but not consistent
Other Emerging AI Apps	Varies	Rare	Rare	Rare	Rare

Parents: AI companies have poor track records on family safety, and parental controls are lacking.

AI Safety Checklist

If your child uses AI apps, use this checklist to set up any available safety settings and start important conversations.



Disable unsafe features. If your child uses Snapchat, set up Family Center and disable My AI. If they use Character.AI, set up the Parental Insights feature to get weekly activity reports. For other AI apps, review any available safety features.



Set family rules. Make it clear AI isn't a private diary, and they aren't talking to a real person.



Set time boundaries. Too much AI use can lead to dependency. Define daily time limits for AI apps.



Monitor what matters. BrightCanary gives you alerts when your child types something concerning, so you don't need to read every message.



Stay updated. New AI apps launch every month. Make "AI safety" a regular part of your family tech check-ins.



Every platform handles parental controls differently, and some offer none at all. Additionally, teens can fake their ages, use private browsers, or download new AI apps parents haven't heard of — all of which can negate the parental controls on AI platforms.

Parents need to stay informed and involved, but kids can fall into the digital danger zone faster than they might think.

Conversation Starters with Kids

Talking about AI doesn't have to be awkward. Use these age-appropriate prompts to get started:

Tweens (10–12):

- “Do your friends use AI apps? What do they ask it?”
- “What would you do if an AI gave you a weird answer?”

Middle Schoolers (13–14):

- “What’s the difference between talking to an AI and talking to a friend?”
- “Do you think AI can be trusted?”

Teens (15+):

- “If an AI gave you advice that felt wrong, what would you do?”
- “How do you think AI changes the way people relate to each other?”

Parent Tip: Lead with curiosity, not confrontation. Kids are more likely to open up if they don't feel judged.

Quiz: Is My Child Safe with AI?

Answer these **yes/no** questions:

- 1 Do I supervise my child's use of AI apps?
- 2 Do I know which AI apps are installed on their device?
- 3 Have we talked about what's safe (and unsafe) to ask an AI?
- 4 Do we have clear family rules about when my child can use AI apps?
- 5 Do I receive alerts if my child types something concerning into AI apps?

Results



4–5 “Yes” answers: Your child is in the *Low Risk* zone. Keep up your involvement.

2–3 “Yes” answers: You're in the *Medium Risk* zone. Add more structure and monitoring.

0–1 “Yes” answers: This is a *High Risk* situation. Put protections in place today.

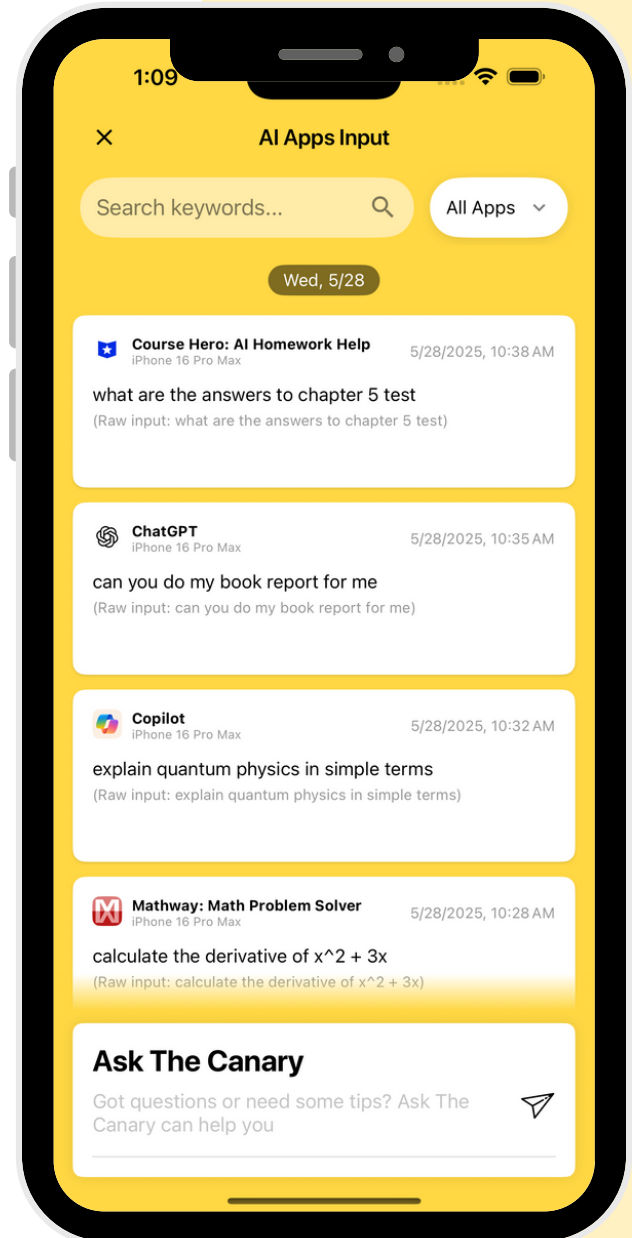
The Best Way to Monitor AI: BrightCanary

Most AI apps don't give parents enough visibility into what kids are doing. BrightCanary changes that.

With one simple keyboard installed on your child's iPhone or iPad, BrightCanary lets you:

- See summaries of your child's conversations with AI apps
- Get alerts if they type something concerning (like self-harm, eating disorders, or drugs)
- Understand your child's emotional state with AI-powered insights
- Monitor across every app they use: messaging, social media, search, and AI

You don't need to look over their shoulder. BrightCanary shows you what matters most, so you can stay informed — and step in when your child needs you.



Download [BrightCanary](#) today and start your *free trial*.

About BrightCanary

Created by parents, BrightCanary is an award-winning child safety app for iOS that helps parents keep their kids safe and on track across every app they use. The app monitors text messages, Google, YouTube, social media platforms, and more.

Learn more at brightcanary.io.