



The Family AI Safety Guide

Settings, scripts, and warning signs every parent of an 8-to-12-year-old needs to know.

BY PAX EMBER

paxember.com

WELCOME

You're here before the day something goes wrong.

Before you read another word, I want to tell you how I think about AI safety with kids.

Most parents discover they need a safety guide the day something goes wrong. You're here before that day. That already puts you ahead of most families.

What you're holding isn't a lecture. It's a working document. Short on purpose. Built to be printed, marked up, taped to a fridge, updated every six months when platforms change again.

You don't need to become an AI expert. You need to know the four things that actually protect your child.

1. Platform settings you should change today
2. Conversations you should have this week
3. Warning signs you should recognize
4. How to tell when AI made something that looks real

Nothing more. Nothing less.

Let's begin.

Pax Ember

hello@paxember.com



SECTION ONE

Platform Privacy Settings

Before you talk to your kid about anything, close the easiest doors first. The settings below are non-controversial, reversible, and most parents haven't touched them. Ten minutes of work here is worth more than an hour-long lecture later.

YouTube (and YouTube Kids)

RECOMMENDED FOR KIDS 8-12: YOUTUBE KIDS OR SUPERVISED ACCOUNTS

YouTube Kids is designed for ages 4 to 12 and has parental controls baked in. Supervised Accounts are the bridge for kids under 13 who want more content than YouTube Kids offers.

SETTINGS TO CHANGE TODAY

- **Content level.** In YouTube Kids parent settings, set to “Younger” (ages 5-8) or “Older” (ages 9-12). “Approve content yourself” is the strictest option and worth considering for kids under 10.
- **Turn off Search.** Kids can still watch videos you've approved without needing to search.
- **Disable autoplay.** Stops the algorithm from pulling your kid into a rabbit hole after the video they came for.
- **Daily Shorts limit.** As of March 2026, parents can set the daily Shorts feed limit to zero. Shorts are the most addictive format on YouTube. Consider setting to 0 for kids under 12.
- **Clear watch history weekly.** Stops the algorithm from over-targeting your child based on one-off curiosity clicks.

WHERE TO FIND THESE

Open YouTube Kids on your child's device, tap the lock icon, enter your parent PIN, Settings, choose the profile.

TikTok (Family Pairing)

Honest note. TikTok's minimum age is 13. Kids under 13 should not have regular TikTok accounts. The under-13 experience is a separate, US-only product with very limited features.

If your child (age 13 or older) uses TikTok, Family Pairing lets you connect your account to theirs and control 15 or more settings from your phone.

SETTINGS TO ENABLE VIA FAMILY PAIRING

- **Private account by default** (teens 13-15 get this automatically in 2026, but verify).
- **Restricted Mode** filters mature content.
- **Screen time limit.** Default is 60 minutes for teens. Set lower for 13 to 15 year olds.
- **Direct messaging off** for ages 13 to 15 (default in 2026, verify it stayed that way).
- **Who can comment on their videos.** Set to "Friends only" or "No one" for younger teens.
- **Download restrictions.** Prevent their videos from being saved by strangers.

WHERE TO FIND THIS

On your phone, open TikTok, Profile, Menu, Settings and privacy, Family Pairing, Continue as a parent, Link accounts.

ChatGPT (OpenAI Parental Controls)

MINIMUM AGE 13. PARENTAL CONTROLS ARE FOR ACCOUNTS AGES 13 TO 17.

OpenAI rolled out actual parental controls in late 2025 after significant pressure from parents and regulators. If your teen is using ChatGPT, these are not optional to know about.

SETTINGS TO ENABLE BY LINKING YOUR ACCOUNT TO THEIRS

- **Reduced graphic content.** Filters out disturbing imagery and language.
- **Blocked content types.** Viral challenges, sexual or romantic roleplay, violent roleplay, extreme beauty ideals.
- **Voice mode.** Can be turned off. Voice conversations can be more absorbing than text for some kids.
- **Image generation.** Can be restricted.
- **Memory.** Can be cleared or turned off entirely. Controls how much ChatGPT "remembers" about your child across sessions.
- **Quiet hours.** Times when ChatGPT cannot be used. Set these for bedtime, school hours, family time.
- **Safety notifications.** You get an SMS, email, or push notification if OpenAI's system detects conversations that may indicate serious self-harm risk.

What you don't get. Access to your teen's actual conversations. Privacy is preserved unless the system flags a serious risk.

WHERE TO FIND THIS

Sign into ChatGPT, click your profile, Parental Controls, Link a teen account. Your teen gets an invitation email, accepts, and the link activates.

Roblox (Kids, Select, and Age-Based Accounts)

Roblox is enormous for kids 8 to 12 and most parents don't realize how much AI is in it. Chat moderation, game recommendations, generative asset creation. It's all AI under the hood.

2026 update. Roblox introduced age-based accounts in June 2026.

- **Roblox Kids** (ages 5-8): chat disabled entirely, limited game access.
- **Roblox Select** (ages 9-15): limited chat, parent can approve specific games.
- **Standard account** (16+): full access.

SETTINGS TO VERIFY ON YOUR CHILD'S ACCOUNT

- **Age set correctly.** If your child set their own age wrong during signup, chat and content access won't match what they actually need. Reset it in their account settings.
- **Chat set to "Friends only" or "No one."** Under 9, chat should be off entirely. Ages 9 to 15, friends-only at minimum.
- **Facial age checks.** Roblox now requires facial age verification to access any chat features. Ensure your child hasn't bypassed this by using a parent's face.
- **Block specific games.** Parents can block individual games through age 15. Use this for any game you've seen controversy about.
- **Screen time limit.** Set in Family Settings. Default is unlimited, which is a problem for most 8 to 12 year olds.

WHERE TO FIND THIS

Roblox app, Settings, Parental Controls, enter 4-digit PIN, Account Restrictions.

YOUR CHECKLIST

The AI Safety Checklist

Print this. Post it on the fridge. Check each box when done.

YOUTUBE

- ☐ YouTube Kids or Supervised Account set up
- ☐ Content level set to age-appropriate
- ☐ Search off (for younger kids)
- ☐ Autoplay off
- ☐ Shorts limit set to 0 (or low)

TIKTOK (if applicable, age 13+)

- ☐ Family Pairing linked
- ☐ Account set to private
- ☐ Restricted Mode on
- ☐ Screen time limit set
- ☐ Direct messaging off

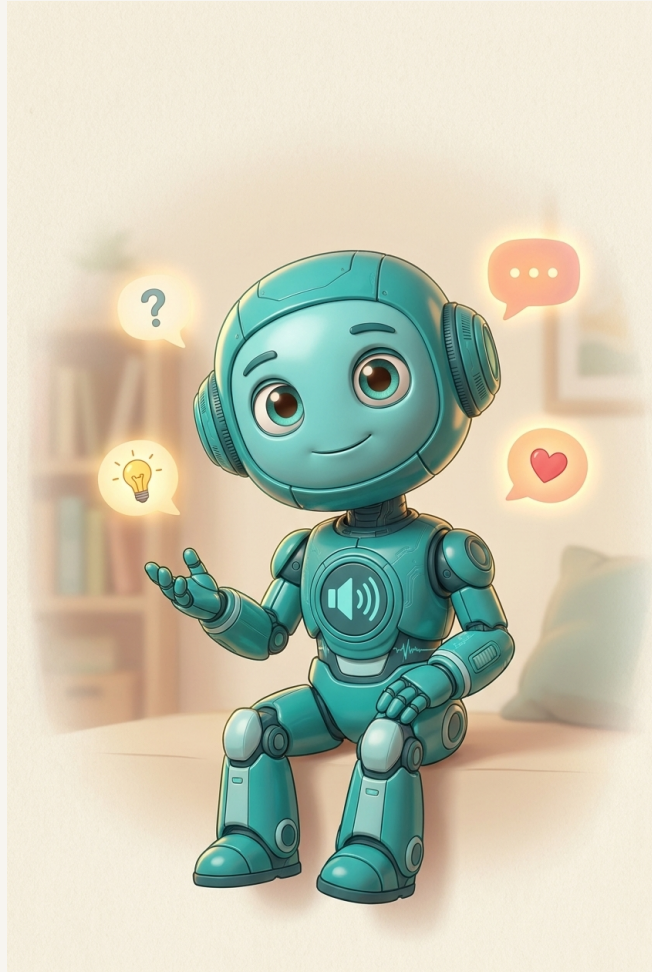
CHATGPT (if applicable, age 13+)

- ☐ Parental controls linked
- ☐ Content filters on
- ☐ Memory cleared
- ☐ Quiet hours set
- ☐ Safety notifications active

ROBLOX

- ☐ Age set correctly
- ☐ Account type matches age (Kids or Select)
- ☐ Chat restrictions set
- ☐ Screen time limit set
- ☐ High-risk games blocked

Review every 6 months. Platforms change their defaults. Your settings don't update themselves.



SECTION TWO

Conversation Starters by Age

Settings protect your child in the background. Conversations shape who they become. The questions below are the ones I would start with. Pick one per week. Don't force the answer.

The silence between the question and the answer is where the
real thinking happens.

Starting points for younger kids

These aren't trick questions. They're real starting points. Pick one per week. Don't force the answer.

ABOUT USING AI

1. "Show me the last thing AI did for you today."

Most kids say they haven't used AI, then realize they've used it twenty times before breakfast.

2. "Have you ever asked ChatGPT or Alexa or Google for something? What did you ask?"

ABOUT UNDERSTANDING AI

3. "Can you explain what AI is, in your own words?"

4. "Where do you think AI learns from?"

Good answer: examples from the internet. Helps them understand it's not magic.

5. "What's something you could do that even the smartest AI couldn't?"

Deeper questions for older kids

ABOUT AI AND HONESTY

6. "If someone made a video using AI that looked like your teacher saying something weird, how would you know it wasn't real?"

7. "Is it cheating to use AI to help with homework? Why or why not?"

The answer they give tells you everything about their school's culture and their own ethics.

ABOUT AI AND RELATIONSHIPS

8. "Have you ever used an AI chat app like a friend? To talk to when you were bored or upset?"

This is the most important question in this whole guide. AI companions are marketed to kids and can be unhealthy. Ask gently. Listen carefully.

9. "If an AI bot acted like your friend, would that feel weird to you? Why or why not?"

ABOUT THE FUTURE

10. "What's one thing you hope AI can do someday that it can't do yet?"

Pure curiosity question. Ends the conversation on a good note.



SECTION THREE

Red Flags to Watch For

None of what follows is proof of harm. All of it deserves a second look. The goal isn't to panic. It's to know what's in front of you, so you don't mistake a phase for a real problem, or a real problem for a phase.

Red flags in your child's AI use

If you notice any of these, it's worth a gentle conversation. None are automatic crises. All deserve attention.

1. Extended time with an AI chat app.

Apps like Character.AI, Replika, and Snap My AI are designed for extended engagement. Hours-long sessions with any chat bot is worth asking about.

2. Your child refers to an AI as “they” or by name, like a person.

“Sage said something funny today.” This anthropomorphizing is a warning sign. AI isn't designed to love them back. It's designed to keep them engaged.

3. Your child hides what they're doing on their device.

Not a conversation about privacy. A conversation about what changed that they now feel it needs to be hidden.

4. Emotional changes tied to device use.

If your child comes off a device and is noticeably different, upset, withdrawn, or agitated, the content matters more than the screen time.

5. Your child defends AI decisions they don't understand.

“The algorithm chose this video, so it must be what I want.” Shows they've stopped seeing AI as a tool and started seeing it as an authority.

Red flags in AI content they encounter

6. Too-perfect photos of “real” kids or events.

Skin looks airbrushed. Eyes look slightly off. Text in the background is nonsense. That's AI-generated.

7. A “friend” they only talk to online who seems to know too much.

Could be an AI bot, could be a scammer using AI tools to seem human. Either way, not safe.

8. Videos of celebrities or politicians saying shocking things.

Deepfakes are cheap in 2026. If it sounds wildly out of character, it probably isn't real.

9. Homework that suddenly improved overnight.

Not a punishment conversation. A teaching conversation. “AI is a great collaborator when you're driving. When it's driving, you stop learning.”

10. New “friend” from a game or app asking personal questions.

Always suspicious. Always flag to parent. Always report.



SECTION FOUR

Verify What You See

This is the skill most parents are intimidated by. It shouldn't be. In five minutes, any family can learn to spot most AI-generated content. Teach your kid these techniques once, and they'll do them automatically for the rest of their lives.

Quick verification moves any family can use

1. The reverse image search move.

Right-click any suspicious image, "Search image with Google." If it shows up on hundreds of sites with different captions, or only on social media with no original source, it's likely fake or stolen.

2. The "look at the hands" test.

AI still struggles with hands. Count the fingers. Check the proportions. Check where fingers meet the palm. Off fingers equals AI image.

3. The "read the text in the background" test.

AI often generates gibberish text in signs, documents, book spines. Zoom in. If the text looks like broken letters or nonsense words, the image is AI-generated.

4. The "who said it first" test.

For quotes or news claims, search the exact phrase in quotation marks. If the only sources are social media posts or AI-generated articles, it's not real news.

5. The "smell test" test.

The most viral AI content is designed to trigger outrage or awe. If a claim feels engineered to make you share it before you verify, that's the signal. Slow down. Check.

Teaching your kid to verify

The skill you're teaching isn't "don't trust anything." It's "know how to check."

ONE RULE TO TEACH

"Before you share something online, ask: would this change what someone thinks about the world? If yes, verify it."

That's the whole skill. The verification techniques on the previous page are the tools. The rule is the habit.

Practice together

Next time you see a wild post online (yours or theirs), stop and verify it together. Use the reverse image search. Check the hands. Read the background text. Find the original source.

After three or four times, your kid will start doing it automatically. After a dozen, they'll be the one teaching their friends.



What comes next

This guide is a starting point. Here's what actually matters going forward.

You don't need to get this perfect. You need to keep the conversation going. The families who do this well aren't the ones with the most locked-down devices. They're the ones where the parent and the kid talk about what they're seeing, what they're using, what they're feeling.

Revisit this guide every 6 months. Platforms change settings. New AI apps appear. Your child gets older. What worked at 9 won't work at 11.

If you want more, three places to go:

→ **The Pax Ember Blog** at paxember.com/blog for essays on AI and parenting, updated regularly.

→ **Welcome to the AI Generation (Book 1)** is the trilogy my readers are using to learn AI alongside their kids. paxember.com/books

→ **Reply to any email from me** at hello@paxember.com. I read them. It's a small list, and I like keeping it that way.

Thank you for showing up. That's the whole thing.

Pax Ember

hello@paxember.com | paxember.com