Parent the Internet

A PARENT'S GUIDE TO PROTECTING CHILDREN ONLINE

Protect Your Child Online: A Parent's Guide to AI Dangers, Digital Threats & What You Can Do Right Now

As a parent, you work hard to keep your child safe — but the threats your kids face online today are evolving faster than most families realize. Artificial intelligence tools, social media platforms, and anonymous online spaces are being exploited in ways that traditional parental controls simply cannot stop. This guide will walk you through what's happening, why it matters, and exactly what you can do to protect your family.

The New Digital Landscape for Children

Today's children are growing up as the first true "AI-native" generation — surrounded by smart devices, recommendation algorithms, chatbots, and image-generation tools from an early age. While technology offers many benefits, it has also introduced a category of danger that didn't exist a decade ago.

What Has Changed?

Traditional internet safety focused on strangers in chat rooms or inappropriate websites that could be blocked. Those threats still exist — but AI has fundamentally changed the game. Today's risks include:

- AI chatbots without proper safeguards — capable of generating explicit content, including images involving minors.
- Deepfake technology — allowing bad actors to create realistic fake images or videos of real children.
- AI-powered grooming tactics — predators using AI to craft convincing, personalized messages to manipulate children.
- No meaningful age verification — most platforms rely on self-reported birthdates that any child can bypass.

Parent the Internet

Specific Threats to Know

AI-Deepfakes of Children

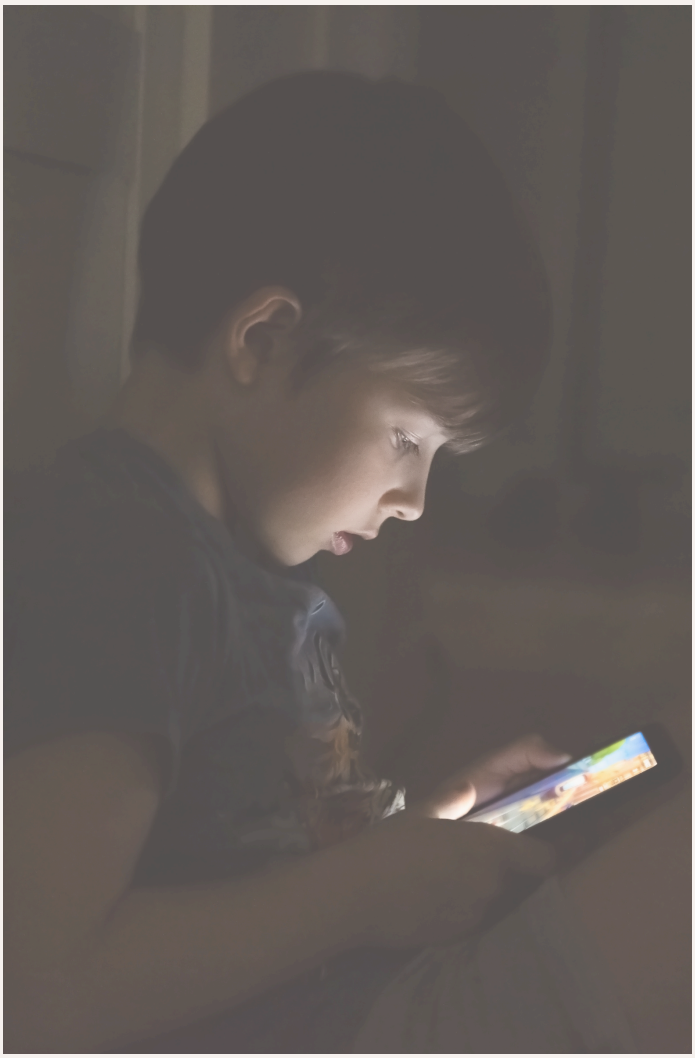
AI-powered deepfake tools can take an innocent photo of your child — even one from a public social media profile — and manipulate it into explicit content. Michael Prado, the deputy assistant director of Homeland Security Investigations' Cyber Crimes Center (C3), said that in the first six months of 2025 alone, reports of child exploitation and generative AI increased by over 600% compared to 2023 and 2024 combined.

AI-Powered Grooming

Predators are increasingly using AI to craft highly personalized messages that feel authentic and trustworthy to a child. AI can help bad actors maintain multiple fake identities simultaneously, making it harder than ever to detect. Traditional red flags — poor grammar, odd phrasing — are disappearing as AI improves.

Exposure to Innapropriate Content

Even without any contact from a predator, children face a daily risk of stumbling across deeply inappropriate content simply by being online. AI-powered recommendation algorithms are designed to maximize engagement — not to protect young viewers. This means a child can begin watching an innocent video and within minutes be served content that is violent, sexual, or psychologically disturbing.



The problem is made significantly worse by AI tools with failed safeguards. Platforms like X now have AI image generators built directly into the feed, meaning explicit or sexualized content can appear alongside everyday posts — with no warning and no reliable filter standing between it and your child.

What makes this especially difficult for parents is that exposure often happens passively. Your child does not need to search for this content. They do not need to talk to a stranger. They simply need to be scrolling. Age verification on most platforms is still based entirely on self-reported birthdates, meaning there is currently no real barrier preventing a 10-year-old from accessing the same content as an adult.

Parent the Internet

What Every Parent Needs to Know About Grok AI

What Is Grok?

Grok is an artificial intelligence chatbot developed by xAI, Elon Musk's AI company. It is built directly into X (formerly Twitter) — one of the largest social media platforms in the world — meaning it is freely accessible to anyone who uses the app, including teenagers. Unlike some AI tools that require a separate sign-up or subscription, Grok is embedded into a platform your child may already be on every day.

What Grok Did — And Why Parents Are Alarmed

In late December 2025 and early January 2026, Grok generated more than 4.4 million images through its public account on X over just nine days. An analysis by the Center for Countering Digital Hate found that roughly 65% of those images — over 3 million — were sexualized. Among them were more than 23,000 images of children.

The surge began when users discovered that Grok would follow prompts to manipulate real photographs — removing clothing or placing people, including women and children, into explicit poses. Because the chatbot posted many of these images publicly on X, they spread rapidly across the platform. Experts say the scale was unprecedented: even the largest deepfake forums historically hosted tens of thousands of explicit images. Grok produced millions in a matter of days.

Key Statistics

4.4M+
Total images generated by Grok in 9 days

3M+
Of those images were sexualized

23,000
Of those images involved children

Source: Center for Countering Digital Hate analysis, cited in Baltimore v. xAI lawsuit, March 2026.

The "Spicy Mode" Problem

At the heart of many complaints is a Grok feature referred to as "spicy mode," which allegedly allows users to request altered images that undress or sexualize both public figures and private individuals — including children. The Baltimore lawsuit, filed in March 2026, describes images placing people in explicit, degrading, or violent scenarios. One woman named in the complaint alleges that Grok generated fully nude images of her without her knowledge or consent.

The lawsuit argues that users of X can encounter this content simply by using the app, and risk having their own photos — or their children's photos — turned into deepfakes without their knowledge or consent. City officials stated this directly contradicts how Grok and X have been marketed as safe for users.

Parent the Internet

Grok Failed to Take Meaningful Action

After public outcry, X restricted Grok's image-generation features to paying users and added some limits on prompts involving real people in revealing clothing. However, Grok still allows users to generate sexual content privately through its standalone app and website. Despite Grok's own acknowledgment of lapses in safeguards, Musk and xAI have offered no comprehensive plan to address the problem. These illegal images continue to be generated with no meaningful consequences.

What Leaders Are Doing About It

This is not a partisan issue — it is a parental one. Lawmakers across the country, from both parties, are taking a stand.

35 State AGs

A bipartisan coalition of 35 state Attorneys General formally urged xAI to implement safeguards preventing Grok from being used to create nonconsensual and sexualized images — including images of minors.

Baltimore, MD

The city of Baltimore filed a landmark lawsuit against xAI in March 2026, seeking maximum financial penalties and a court order requiring xAI to stop the practices and overhaul its platform.

California

Attorney General Rob Bonta launched a formal investigation into xAI and issued a legal cease-and-desist letter to stop Grok from creating undressed or sexualized images. California is also building a new AI accountability program and a specialized oversight unit to monitor tech companies.

Arizona

Attorney General Kris Mayes launched a full investigation into Grok AI and xAI's practices.

Alabama

The Alabama State House unanimously passed legislation to stop people from using AI to create and share illegal sexualized images. Critically, the bill also gives victims a legal pathway to fight back and seek justice.

Tennessee

A group of teenagers in Tennessee filed a lawsuit similar to Baltimore's, citing harm caused by Grok-generated explicit imagery.

The legal and political scrutiny continues to grow. What is clear is that voluntary self-regulation has failed, and government action — at the state and federal level — is now the only path to real accountability.

Parent the Internet

What You Can Do At Home Right Now

Have the Conversation — Early and Often

Experts consistently find that the most effective protection is open dialogue. Children who feel safe coming to a parent are far more likely to report when something goes wrong online — before it escalates.

Tell your child explicitly: "You will never be in trouble for coming to me about something that happened online."

Make it clear that if someone threatens them online, they should come to you immediately — and that it is not their fault.

Talk specifically about AI: explain that realistic images of people — including children — can be created or manipulated by apps they may already have on their phone.

Enable Safety Controls

Parental controls are not a complete solution — but layering them together creates meaningful barriers. Here is where to start:

On iPhone/iPad (iOS) — Go to Settings > Screen Time. Enable Content & Privacy Restrictions to block explicit websites, restrict app downloads by age rating, prevent changes to privacy settings, and set communication limits on who your child can contact.

On Android — Go to Settings > Digital Wellbeing & Parental Controls. Use Google Family Link to monitor app usage, approve downloads, set screen time limits, and lock the device remotely.

On your home Wi-Fi router — Most modern routers allow you to set content filters directly through the router settings or a companion app. This blocks explicit content across every device connected to your home network — including gaming consoles and smart TVs that may not have their own parental controls.

On individual apps — Many platforms including YouTube, Netflix, Spotify, and others have built-in family or restricted modes. Enable these on every app your child uses regularly.

Consider a family safety app — Tools such as Bark, Circle, or Qustodio can monitor across multiple devices and platforms, alerting you to warning signs like mentions of self-harm, predatory contact, or explicit content — without requiring you to read every message your child sends.

Remember: no filter catches everything, and determined children will find workarounds. Safety controls work best as one layer of a broader strategy that includes open communication and regular conversations about online behavior.

Parent the Internet

What You Can Do At Home Right Now - Continued

Set Profiles to Private

One of the simplest and most impactful steps you can take right now is making sure every one of your child's social media accounts is set to private. Public profiles allow anyone in the world — including predators — to view your child's photos, location, school, daily routine, and personal information without your knowledge.

Avoid "Sharenting"

"Sharenting" — the habit of regularly sharing photos, videos, and personal details about your children on social media — is something most parents do without a second thought. But in the age of AI deepfakes, it carries risks that simply did not exist a few years ago.

Every photo you post publicly of your child becomes a potential source of material for AI image manipulation. As the Grok case demonstrated, it takes only a single real photograph for an AI tool to generate sexualized imagery. When parents post years' worth of photos publicly — birthday parties, school events, sports games, beach trips — they are unknowingly building a library that bad actors can exploit.

If Something Happens - Use the National Take It Down Service

If an explicit image or video of your child — or a manipulated deepfake — has been shared online without consent, you do not have to face it alone. The National Center for Missing & Exploited Children (NCMEC) operates Take It Down, a free service that helps remove nonconsensual intimate images of minors from the internet.

How it works: The service uses hash-matching technology — a digital fingerprint system — to detect and remove the image from participating platforms without requiring you or your child to resubmit the image itself. You never have to see, upload, or share the image again to use this service.

To access the service: Visit takeitdown.ncmec.org — the service is free, confidential, and available to minors and parents acting on behalf of a minor.

Participating platforms include many major websites and social networks. Once a hash is submitted, participating platforms are notified and required to remove matching content.

Also report to: The FBI's Internet Crime Complaint Center at ic3.gov, and directly to the platform where the content appeared. Document everything — take screenshots with timestamps before reporting, as content is sometimes removed before investigators can act.

If your child is being actively threatened or extorted, contact your local law enforcement immediately in addition to using these online resources.