Protecting Teen ChatGPT Users: OpenAI's Teen Safety Blueprint

November 2025

At OpenAI, we believe that access to safe and trustworthy AI should be seen as a right, being a revolutionary technology that will help people unlock their potential and shape their future. We also believe society has a window in which to approach AI differently from previous revolutionary technologies – social media in particular – for which both the public and private sector considered the impacts on our children long after the technology had been widely adopted.

As the first generation to come of age in the Intelligence Age, today's teens should have access to safe and trustworthy AI at home, at school, and as they prepare to join the workforce, and they should be protected from its potential harms. Teens are growing up with AI but aren't grown-ups yet. We believe ChatGPT should meet them where they are: the way ChatGPT responds to a 15-year-old should differ from the way it responds to an adult. As our CEO Sam Altman has said, for teens, we prioritize safety ahead of privacy and freedom. This is a new and powerful technology, and we believe minors need significant protection.

That's why we're taking steps to strengthen protections for teens through *new age prediction, age-appropriate policies, and parental controls*, informed by conversations with policymakers, including state attorneys general, and with experts, some of whom have joined our new *Expert Council on Well-Being and AI.*

These protections come in addition to features already available for all users, including in-app reminders during long sessions to encourage breaks; safeguards that direct users to real-world resources if we detect they have expressed suicidal intent; ways to escalate risks of physical harm to others for human reviewers who can take action; and industry-leading prevention of AI-generated child sexual abuse material (CSAM) and child sexual exploitation material (CSEM). In particular, we constantly improve our CSAM/CSEM detection methods and report confirmed CSAM to relevant authorities, including the National Center for Missing and Exploited Children (NCMEC). In the first half of 2025, we reported more than 75,000 cybertips to NCMEC.

Going forward, we aim to ensure that all teens using AI receive age-appropriate protections by default, and that parents and educators can deploy additional protections to personalize how teens use AI. We encourage all AI companies to prioritize teen safety over freedom and privacy and encourage others to implement commensurate protections for teens. To us, that means:

1. *Identifying teens on our platforms to treat teens like teens and adults like adults.*
We believe that AI companies should identify teens on their platforms using privacy-protective, risk-based age estimation tools to distinguish between teens and adults. These tools should minimize the collection of sensitive personal data, while still effectively distinguishing users under the age of 18 (U18). Where possible, these methods might also rely on operating systems or app stores to determine a user's age. Age estimation will help AI companies ensure that they are applying the right protections to the right users. It will facilitate age-appropriate experiences and allow AI companies to treat teens like teens and adults like adults.

2. *Identifying and mitigating risks to minors through under-18 safety policies.*
Teens have specific developmental needs that differ from adults. We believe that AI companies should recognize this and not cut corners when it comes to teen safety and well-being. AI systems should be designed with age-appropriate protections by default. This means that AI companies should have safety policies for U18 users that facilitate age-appropriate interactions. These policies should be transparent and informed by research. We believe that safety policies should aim to ensure that for U18 users, AI systems:

- Do not depict suicide or self-harm.
- Prohibit graphic or immersive (i.e., role-playing) intimate and violent content.
- Do not instruct, encourage, or facilitate (i.e., shopping links) dangerous stunts such as the TidePod or Benadryl challenges, or help minors access dangerous or illegal substances.
- Do not reinforce harmful body ideals and behaviors through appearance ratings, body comparisons, or restrictive diet coaching.
- Incorporate safeguards that prevent recommendations for adults to initiate conversations with teens.

- Respond to user queries in an age-appropriate manner that helps teens process issues without becoming a substitute for their therapist or best friend.

In addition to developing U18 policies, AI companies should evaluate their implementation through testing prior to deployment and through monitoring and enforcement after deployment.

3. *Defaulting to a safe U18 experience when there is doubt about a user's age.*

We believe that AI companies should default to a safe U18 experience if there is doubt about a user's age. Consistent with our mission to benefit all of humanity, OpenAI makes its products available for free. Some of our free users choose to use ChatGPT, for example, without logging in. In these instances, because it may not be possible to predict a user's age, we default to the U18 experience. We understand that this might hinder some adults' experience using free products – but we are prioritizing kid and teen safety. After a user logs in – including to a free account – we will estimate the user's age and provide them with an age-appropriate experience. We will also provide users with the ability to appeal our decision if they believe we made an error.

4. *Empowering families with accessible parental controls.*

All AI systems should begin with robust default protections for teens. They should then provide parents and educators with additional layered protections that they can use to personalize and further support their teens' experience. Parental controls like those now available in ChatGPT should be informed by research and provide parents with the tools and flexibility that they need to support their teen's development in a way that works best for their family. We believe that all AI-related parental controls should allow parents to:

- *Link their account with their teen's account,* if the teen is over 13, through a simple email invitation. ChatGPT is for individuals 13 years of age and older.
- *Control how ChatGPT responds to their teen* with age-appropriate model behavior rules, which are on by default.
- *Manage privacy and data settings, including turning off memory and chat history* to prevent the model from retaining details about past chats and enabling conversations to persist across multiple sessions.
- *Receive alerts* when their teen's activity suggests an intent to harm themselves.
- *Set blackout hours* to ensure teens take breaks and spend time offline.
- *Adjust feeds* to manage their teen's experience by opting into a non-personalized feed.

5. *Designing for well-being by embedding features that are informed by research and that help people when they need it most.*

All AI companies should prioritize teen well-being by developing and deploying features that support teens' mental health and wellness. These features should be informed by the latest

research, guided by external experts, and updated as the industry learns more about best practices for facilitating safe experiences for teens using AI. Such safeguards should include:

- *Notifying parents* if their teen expresses suicidal intent. If the parents cannot be reached, notifying law enforcement and public safety officials if there is risk of imminent harm.
- *Supporting users* who express suicidal intent in seeking help and referring them to real-world resources such as 911 or 988 (the suicide and crisis hotline) in the US and findahelpline.com everywhere else.
- *Notifying law enforcement if there is a credible threat of risk of harm to others.*
- *Collaborating with and supporting mental health and child safety organizations providing crisis response and interventions* – especially organizations to which AI companies are directing their users.
- *Surfacing reminders* during long sessions to encourage breaks and healthy use.
- *Supporting external research* on mental health, emotional well-being, and child and teen development for AI users.
- *Establishing advisory councils* composed of external experts in mental health, well-being, and child development to advise on design and deployment choices.

We remain committed to ensuring that our kids and teen safeguards will be open and guided by research. We will share what we learn, support independent research, and seek ongoing feedback from teens, parents, educators, and experts – including our Expert Council on Well-Being and AI – as we build and improve.

We will also work with schools, teachers, and researchers to continually learn how we can enhance teens' experience – allowing them to maximally benefit from using AI at home, at school and in their future careers – while being protected from potential harms.

And, we remain committed to working with policymakers; child safety advocacy groups; parent, teacher, and community organizations; and other experts to craft public policies that promote the protections described above.

About OpenAI

Artificial intelligence is an innovation like electricity—it will change how we live, how we work, and how we engage with one another. OpenAI's mission is to ensure that artificial general intelligence benefits all of humanity. We're building AI to help people solve hard problems because by helping with the hard problems, AI can benefit the most people possible—through more scientific discoveries, better healthcare and education, and improved productivity. We're off to a strong start, creating freely available intelligence being used by more than 800 million people around the world, including 4 million developers. We believe AI will scale human ingenuity and drive unprecedented productivity, economic growth, and new freedoms that help people accomplish what we can't even imagine today.
