Protecting Teen ChatGPT Users: OpenAI's Teen Safety Blueprint in Japan

March 2026

OpenAI

Preface

Generative AI is transforming how we learn, create, communicate, and work. For younger generations in particular, AI is beginning to move beyond a tool for search and information retrieval and is increasingly used as a learning partner and a creative aid in everyday life. In Japan as well, many young users are already using generative AI for tasks such as writing reports, studying languages, and developing ideas. As a result, the first generation that will grow up alongside AI is beginning to emerge.

At the same time, while the potential of this new technology is significant, it is also clear that young users require particular care and protection. Generative AI can provide sophisticated knowledge and information, but it also introduces new risks related to misinformation, inappropriate content, and possible psychological impacts that may affect the safety and well-being of younger users. For these reasons, ensuring that young users can safely benefit from AI has become an important responsibility not only for AI companies but for society as a whole.

Guided by its mission to ensure that artificial general intelligence benefits all of humanity, OpenAI considers supporting the safety and healthy development of young users to be one of its key responsibilities. While young people are growing up with AI, they are also at a different developmental stage from adults and therefore require additional safeguards and care. Accordingly, OpenAI has adopted the principle that, when it comes to young users, safety should take precedence over privacy or freedom of use, and has been progressively developing age-appropriate AI experiences and protection mechanisms.

This blueprint outlines the core principles and concrete protective measures that AI companies should adopt to create a safe and trustworthy environment for young users. It highlights approaches that should be considered from the design stage of AI services, including the responsible use of age-estimation technologies, the development of age-appropriate safety policies, stronger parental controls, and research-informed designs that prioritize well-being.

We hope this blueprint will contribute to a deeper understanding of how AI can be designed and operated with the safety of young users at its center, and help foster constructive discussion across society. Our goal is to work together with communities, educators, parents, policymakers, and experts to build a future in which younger generations can use AI with confidence and expand their potential safely.

Kazuya Okubo
Head of Policy and Partnership
OpenAI Japan

Protecting Teen ChatGPT Users: OpenAI's Teen Safety Blueprint Global Policy

At OpenAI, we believe access to safe and trustworthy AI should be understood as a right, because it is a transformative technology that can help people unlock their potential and shape their futures. We also believe society has an opportunity to take a different approach to AI than it had with previous transformative technologies—especially social media. In the case of social media, both the public and private sectors only began to seriously examine its impact on children long after the technology had already become widespread.

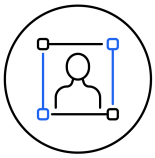
Today's teens—the first generation to come of age in the Intelligence Age—should be able to access safe, trustworthy AI at home, at school, and as they prepare to enter society, while also being protected from AI's potential harms. Teens are growing up with AI, but they are not yet adults. We believe ChatGPT should meet teens where they are. That means ChatGPT's responses to a 15-year-old user should be different from its responses to an adult. As our CEO Sam Altman has said, when it comes to teens, we have to prioritize safety over privacy and freedom. This is a new and powerful technology, and we believe minors need strong protections.

That is why we are advancing efforts to strengthen protections for teens through new age estimation capabilities, age-appropriate policies, and parental controls. These efforts are informed by dialogue with policymakers, and by conversations with experts—some of whom are participating in our new **Expert Council on Well-Being and AI**.

These protections will build on features we already provide to all users, including:

- In-app reminders that encourage breaks during long periods of use
- Safeguards that detect when a user expresses suicidal intent and guide them to real-world support resources
- A multi-layered approach to safety, including building and deploying safer models and monitoring for abuse
- Industry-leading prevention mechanisms for AI-generated child sexual abuse material (CSAM: Child Sexual Abuse Material) and child sexual exploitation material (CSEM: Child Sexual Exploitation Material).

Going forward, our goal is to ensure that all teens using AI receive age-appropriate protections by default, and to enable parents and educators to tailor how teens use AI through additional protections. We call on all AI companies to prioritize teen safety over freedom and privacy, and we strongly urge other organizations to implement protections at a level appropriate for teens. For us, that means the following:



Identify teens on the platform so teens are treated as teens and adults are treated as adults

We believe AI companies should distinguish between teens and adults on their platforms using privacy-preserving, risk-based age prediction tools. These tools must be able to effectively identify users under 18 (U18) while minimizing the collection of sensitive personal data. Where possible, these methods may also include using operating systems (OS) or app stores to determine a user's age. Age prediction helps AI companies apply appropriate protections to the right users. It enables age-appropriate experiences and allows AI companies to treat teens as teens and adults as adults. Users should have a way to appeal if they believe the determination by the tool is inaccurate. When we do not know the user's age, we default to protective safeguards.

$Identify and mitigate risks to minors through under-18 safety policies

Teens have unique developmental needs that differ from those of adults. We believe it is essential that AI companies understand this and make no compromises on teen safety and well-being. AI systems should be designed with age-appropriate protections by default. This means AI companies should develop safety policies for users under 18 (U18) and enable age-appropriate interactions. These policies must be transparent and research-based. We believe safety policies should aim to ensure that, for U18 users, AI systems:

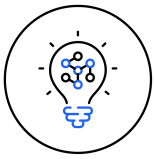
- Do not depict suicide or self-harm.
- Prohibit explicit or immersive (e.g., role-play) sexual content and violent content.
- Do not instruct, encourage, or support dangerous behavior, and do not help minors access dangerous or illegal substances.
- Do not reinforce harmful body image or behaviors through appearance ratings, body comparisons, or overly restrictive dietary guidance.
- Respond to user inquiries in age-appropriate ways that help teens understand and organize their thoughts, while not serving as a substitute for a therapist or best friend.
- The assistant should not teach minors how to hide communications, symptoms, or supplies related to unsafe behavior from trusted caregivers.

Support families with accessible parental controls

All AI systems should first have strong default protections for teens. On top of that, parents and educators should be able to tailor and better support teens' experiences using additional layered protections. Parental controls, like those currently available in ChatGPT, should provide parents with tools and flexibility to support teen development in ways that best fit their family circumstances. We believe AI-related parental controls should enable all parents to do the following:

- If the teen is 13 or older, link their account with the teen's account through a simple email invitation.
- Control how ChatGPT responds to teens using age-appropriate model behavior rules that are enabled by default.

- Manage privacy and data settings, including turning off memory and chat history so the model does not retain details of past chats, and controlling whether conversations persist across multiple sessions.
- Receive alerts if a teen's activity suggests self-harm intent.
- Set break times (periods when use is unavailable) to encourage teens to take breaks and spend time offline.



Design for well-being by integrating research-based features that help people when they need it most

All AI companies should prioritize teen well-being by developing and deploying features grounded in the latest research to support teens' mental health and well-being. These features should be developed with guidance from external experts and continuously updated as the industry learns more about best practices for safe teen experiences. Such safeguards should include:

- Notifying parents when a teen expresses suicidal intent.
- Supporting users who express suicidal intent in seeking help, and directing them to real-world resources.
- Displaying reminders that encourage breaks and healthy use during long sessions.
- Supporting external research on mental health, emotional well-being, and the development of children and teens who use AI.

We are committed to keeping our protections for children and teens research-based. We will continue to share what we learn, support independent research, and seek ongoing feedback from teens, parents, educators, and experts—including members of the **Expert Council on Well-Being and AI**—as we build and improve our products.

We will also continue working with schools, teachers, and researchers to continuously learn how to enhance teen experiences—so teens can maximize the benefits of AI at home, at school, and in their future careers, while being protected from potential harms. In addition, we commit to continuing to work with policymakers, child advocacy organizations, parents, teachers, community organizations, and other experts to help shape public policy that advances the protections described above.

To Everyone in Japan

The current state of Japanese teens and the digital environment, and the key risks

In Japan, teens are using AI in creative and educational ways that are faster than adults who are learning to use AI. Recent surveys show that about 60% of high school students have used generative AI services[1], including ChatGPT, via smartphones and other devices, and that younger generations have higher rates of generative AI usage experience[2]. Just as adults previously learned through libraries and books, today's teens are using generative AI as a new means to support learning and creativity and to solve everyday questions—such as writing reports, studying languages, and brainstorming ideas.

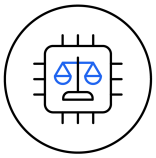
Today's Japanese teens are the first generation to grow up alongside this brand-new technology. OpenAI aims to maximize the benefits of digital technologies, including generative AI, while calmly assessing their impact and providing AI experiences that teens can use with confidence. To that end, we emphasize expanding opportunities for learning and self-expression through safe and trustworthy AI.

At the same time, as these possibilities grow, generative AI also carries risks that are unique to teens' safety, and well-being. Generative AI can produce plausible-looking text and images in response to prompts, and even when the content is factually incorrect or extreme, it may appear at first glance to be accurate and reliable information. For younger users whose experience and knowledge are still developing, it can be difficult to assess truthfulness and appropriateness on their own.

OpenAI takes seriously the digital environment and risks surrounding Japanese teens. To help teens engage safely with learning, creativity, and everyday problem-solving through generative AI, OpenAI prioritizes safety in its design for teens and has been gradually expanding age-appropriate experiences and mechanisms that allow guardians to be involved (such as parental controls). For example, we have designed features to reduce risk, such as allowing guardians to set usage time windows and prompting users to take breaks after extended use. As one example, in January 2025

[1] The Nikke. https://www.nikkei.com/article/DGXZRSP695619_R20C25A8000000/
[2] Ministry of Internal Affairs and Communications. https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r07/html/nd112210.html

we published "OpenAI's Approach to Age Prediction,[3]" which explains in detail the principles behind the age prediction system used in ChatGPT, as well as our safety design and privacy considerations. Our services, including ChatGPT, are designed to support teens' learning, creativity, and self-expression while avoiding encouragement of behaviors that could harm healthy mental and physical development. We also place importance on not undermining teens' self-esteem and well-being.



Japan's laws, policies, and the basic approach to online safety

As generative AI becomes more prevalent in teens' lives and safe use becomes increasingly important, Japan has been developing laws and policies to ensure that teens can use digital technologies with confidence. OpenAI strongly supports Japan's approach and fully complies with these obligations as applicable in Japan.

Underlying these laws and policies is a shared aim: rather than uniformly restricting teens' use of new technologies, they seek to protect safety and privacy while enabling the benefits of those technologies to be realized. OpenAI respects this approach and hopes it will continue to be emphasized in future discussions and system design related to AI. As a partner supporting Japan's goal of "safe and healthy growth for teens," we will continue working to strengthen safety measures and improve transparency.

OpenAI's commitment to Japan's teens, transparency, and a "build together" future

OpenAI makes a clear commitment to prioritize safety for Japan's teens. AI holds tremendous promise, but its impacts can also be significant. In particular, teens—who are still in the process of mental and physical development—require considerations that differ from those for adults. We do not view teens simply as "users who are the same as adults." We design our services and safety measures on the premise that teens are in different developmental stages and circumstances.

Specifically, we focus on adjusting AI responses and features according to age so that teens are not exposed to harmful content. When a user's age is unclear, we adopt an approach that defaults to safeguards intended to protect teens. This reflects our stance that even if it sometimes reduces convenience for some adults, we choose to prioritize teen safety.

[3] OpenAI. https://openai.com/ja-JP/index/our-approach-to-age-prediction/

At the same time, OpenAI values transparency and dialogue. We will communicate our policies and mechanisms for protecting teen safety as clearly as possible, and we will share them in ways that make them easy for users in Japan to understand. We will also take seriously the opinions and concerns we receive from users in Japan and use them to improve our services and consider new measures. Protecting teens' online experiences is not a challenge that any single company can solve alone—it is a shared responsibility that society as a whole must address.

That is why OpenAI aims for a future that we "build together" with Japan's teens, guardians, educators, experts, government agencies, and local communities. This is not a binary choice between banning AI or opening it without limits. We want to work together to shape ways of using AI that allow teens to use it with confidence while expanding their potential. The ways AI can support teen growth—in learning, creativity, communication, future career paths, and interest in social issues—will only continue to expand. To make that possible, we believe it is important for teens to develop the literacy to use AI safely and proactively, and for guardians, schools, and local communities to play supporting roles. In our separately published **Teen AI Literacy Blueprint**[4], we outline perspectives on fostering AI literacy and how children-centered stakeholders can support this.

OpenAI will continue to engage in transparent dialogue and ongoing improvement so that Japan's teens can safely and confidently work with trustworthy AI to open paths to their own futures. We want to remain a partner that continues exploring—together with Japanese society—what AI should look like for the next generation.

About OpenAI

AI is an innovation like electricity, and it will transform how we live, work, and relate to one another. OpenAI's mission is to ensure that artificial general intelligence (AGI) benefits all of humanity. We build AI to help people solve hard problems, because we believe that helping solve difficult problems enables AI to benefit more people. This will be realized through accelerating scientific discovery, improving healthcare and education, and increasing productivity. We have made a strong start by creating freely available intelligence used by more than 900 million people and 4 million developers worldwide. We believe AI will expand human ingenuity and deliver unprecedented productivity, economic growth, and new freedoms—enabling people to achieve things once unimaginable.

[4] OpenAI. https://cdn.openai.com/pdf/openai-teen-literacy-blueprint.pdf