



An Australian Youth Safety Blueprint

September 2026

Foreword

Today's young people will be the first generation to grow up with AI as part of everyday life. Nearly nine in ten teens who use ChatGPT turn to it for learning, information, skill-building, or productivity in a given week. That opportunity comes with a clear responsibility: teens should have experiences designed around their distinct developmental needs, with strong protections by default and meaningful opportunities to learn, create, build skills, and prepare for the future. Getting this right matters.

Unlike with earlier technological shifts, particularly social media, we have an opportunity to put meaningful protections in place from the outset. The responsibility should not fall primarily on young people or their families: companies must build age-appropriate safeguards into their products by default, while giving parents and guardians accessible tools and clear information.

As teenagers embrace AI for learning and creativity, Australia's existing and proposed frameworks—including Australia's proposed Digital Duty of Care reforms, currently under consideration in the recently released draft of the Online Safety Amendment (Digital Duty of Care) Bill 2026 —should provide targeted, harmonised protections that reflect the distinct characteristics of AI while preserving young people's access to beneficial tools. Just as access to books unlocks the value of a library, access to safe AI unlocks new opportunities for personalised learning, helps reduce barriers for the underserved, and expands economic opportunity.

Around the world, governments, educators, researchers, civil society, and industry are working to expand safe access, study real-world outcomes, and develop trusted evidence and practical guidance. Australia can help lead that effort by advancing standards that protect young people while promoting AI literacy, learning, creativity, and future opportunity.

At OpenAI, four commitments guide our work for teens

- **Put teen safety first:** We've built age-appropriate safeguards into the ChatGPT experience.
- **Encourage access to real-world support:** Reinforce trusted relationships and encourage support from people who are in the teen's life.
- **Treat teens like teens:** Reflect their developmental stage without talking down to them, or treating them like adults.
- **Be transparent:** Set clear expectations about how our systems work and how they should behave.

OpenAI has put these commitments into practice through our products, global policy principles, and leadership on advancing strong legislative frameworks, like the Parents & Kids Safe AI Act. In ChatGPT, that includes parental controls to help families guide their teens' use, Under-18 Principles in our Model Spec (as outlined above), and age prediction designed to apply the right protections to the right users. In August we began rolling out ChatGPT for Teens, the new default experience for users identified as ages 13 through 17 in Australia. This experience builds on OpenAI's existing efforts to help young people learn, grow, and develop study skills, with updated guardrails designed to provide a safer environment.

Youth AI safety requires sustained international cooperation, trusted evidence, and practical guidance that can keep pace with the technology. Governments, researchers, civil society, educators, and industry should work together to share evidence, develop common standards, and translate shared goals into effective safeguards. Australia can play a leading role in that effort, helping raise youth AI safety standards across the region and globally.

This Australian Youth Safety Blueprint is intended to support that leadership. It sets out core principles for protections that are evidence-based, tailored to AI, feasible to implement, and focused on meaningful safety outcomes. We hope it provides a practical foundation for an Australian regulatory approach that is ambitious in protecting young people, clear in its expectations, and capable of supporting innovation that benefits the next generation.

Our Blueprint

Six pillars that policymakers should consider for grounded, technology-aware policies that empowers and protects youth in the AI era.

1. Recognise the positive uses of AI and encourage responsible adoption in education

AI policy for young people should start from a simple premise: AI can unlock human potential and expand how young people learn, create, explore, and prepare for the future. Today's teenagers are the first generation coming of age in the Intelligence Age, and they should have access to AI and as they prepare to enter the workforce. The goal for policymakers should not be to keep young people away from AI, but to ensure they learn to use it safely, critically, creatively, and productively.

That requires treating AI literacy as both an opportunity and a safety strategy. AI systems are meaningfully different from other digital services: tools like ChatGPT for Teens are designed for learning and building, not passive scrolling. Today, nearly 9 in 10 teens on ChatGPT use it for learning, information, skill-building, or productivity in a single week. Teens need education that equips students to critically interpret AI outputs, use AI transparently and effectively, and understand both the capabilities and limitations of the technology. They also need parents, teachers, and communities with the resources to guide responsible use.

Not all digital services are the same, and Australian regulation should reflect how AI tools are actually used and make room for beneficial learning, productivity, accessibility, and creativity use cases to grow responsibly. In schools, for example, adoption should be teacher-led and grounded in strong academics. Teachers are irreplaceable and should set the pace for how AI is introduced in and around the classroom. Generative AI has great potential to assist teaching and learning and reduce the administrative workload in Australian schools, which frees up more time for student engagement. As educators build expertise, AI can also support more personalised instruction, active learning, tutoring, and support for multilingual learners and other student populations, while preserving rigorous instruction in core subjects.

State and territory governments have primary responsibility for primary and secondary school education, including decisions about how to implement national initiatives such as the Australian Curriculum. The Productivity Commission has recommended a national approach to educational technology and AI tools, to enable lessons learned in the early adopting jurisdictions to be incorporated into future investments.

2. Identifying teens through privacy preserving age-assurance to treat teens like teens and adults like adults

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 minimise the collection of sensitive personal data, while still effectively distinguishing users under the age of 18 from adults. 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.

OpenAI identifies under-18 Australian users through the same privacy-protective, risk-based age prediction approach we use globally. Where age cannot be predicted with sufficient confidence, ChatGPT will default to a safer experience.

Users incorrectly placed in the under-18 experience can quickly confirm their age and restore full access by submitting a selfie to Persona, a secure third-party identity-verification service. OpenAI does not see the selfie or any identification submitted, and Persona deletes verification data within seven days. Consistent with the Privacy Act 1988, we will continue to minimise the personal information collected through this process.

3. Identifying and mitigating risks to minors through under-18 safety policies.

Teens have specific developmental needs that differ from adults. AI companies should not cut corners on teen safety and wellbeing. AI systems should be designed with age-appropriate protections by default.

AI companies should have safety policies for users under 18 that facilitate age-appropriate outputs and evaluate their implementation through testing prior to deployment and through monitoring and enforcement after deployment. This includes having clear protocols in place to address serious safety situations including self-harm, exploitation, grooming, sexually exploitative content, and other high-risk interactions. These should include in-service support, referrals to appropriate resources, and timely parental notifications when appropriate. Companies should also design their systems to prevent them from generating content that is unsafe or developmentally inappropriate for teens, including graphic sexual or violent material.

To ensure families and users understand the safeguards and tools AI companies have put into place, AI companies should seek to publish robust child safety policies that describe the company's approach to protecting children and teens, including the safeguards it implements and the parental tools available to families.

4. Protect teens from manipulative and deceptive AI outputs

Teens are still developing emotionally, socially, and cognitively. AI providers should recognise these vulnerabilities and design systems that support young people's wellbeing rather than exploit immaturity, trust, or dependency.

This means moving beyond traditional safety concepts and prescriptive one-size fits all defaults to instead address the design underpinning AI outputs. For example, regulation should focus on the specific proven risks policymakers are seeking to prevent — such as unhealthy emotional

dependency or inappropriate relationship dynamics — and require companies to mitigate those risks directly through methods shown to meaningfully mitigate them.

AI systems should have safeguards in place to prevent outputs that encourage emotional overreliance, compulsive engagement, secrecy from parents or trusted adults, or confusion about whether someone is speaking with a machine or a human.

Systems should avoid initiating, reinforcing, or escalating anthropomorphic behaviours that may mislead children into perceiving the AI as conscious, emotionally sentient, a romantic partner, or a trusted human authority figure.

AI systems used by teens should also include wellbeing-oriented interventions where relevant. For example, systems can encourage breaks during extended sessions, signpost trusted helplines or crisis support where distress is detected, and guide users toward offline support from parents, guardians, teachers, or professionals when appropriate.

5. Developing and deploying a crisis response protocol including embedding features that are informed by research and that help people when they need it most.

AI companies should prioritise teen well-being by developing and deploying documental crisis response protocol, and embedding 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 by default parents who have their account connected to their teen's account if their teen expresses suicidal intent.
- Supporting users who express suicidal intent in seeking help and referring them to real-world resources such as Triple Zero (000) for life-threatening situations, Lifeline (13 11 14), Kids Helpline (1800 55 1800) or 13YARN.
- Collaborating with and supporting mental health and child safety organisations providing crisis response and interventions – especially organisations 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 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.

6. 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 personalise and further support their teens' experience. Parental controls like those available in ChatGPT for Teens 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 13 or older, through a simple email invitation. ChatGPT is for individuals 13 years of age and older.
- Manage privacy and data settings, including those related to memory, chat history, and location.
- Automatically receive alerts when their teen's activity suggests an intent to harm themselves.
- Set quiet hours and study hours to promote educational use at certain times of day, encourage regular breaks, and help teens spend time offline.
- Provide timely notice to a parent connected to a child's account when the child modifies or disables a privacy, safety, or parental control setting that was previously enabled or configured by the parent.

Supporting a Risk-Based, Future-Ready Legislative Framework

The six pillars outlined above should be supported by a clear, risk-based legislative framework that requires AI companies to identify, assess, and mitigate risks of severe harm to teens. Rather than prescribing fixed technical measures, the framework should establish proportionate, outcomes-focused obligations that can evolve alongside emerging evidence and technology.

Australia already has a strong foundation for this approach. eSafety's legally enforceable Codes and Standards address class 1 and class 2 material, including certain AI-generated content, and require relevant providers to assess and mitigate risks associated with AI features. These protections operate alongside the National Classification Scheme as part of Australia's broader online-safety framework. We note that the Australian Government is currently developing a Digital Duty of Care framework for online services which would replace the Code and Standards if enacted, with exposure draft legislation currently out for consultation. OpenAI supports the objective of requiring providers to identify and address risks arising from the design and operation of their services. The final framework should establish clear, proportionate obligations tailored to each digital service's functions, users and risks, including the distinct characteristics of generative AI. It should build coherently on existing online safety codes and standards, allow safeguards to evolve with research and technology, and preserve young people's access to beneficial AI tools.

OpenAI is currently consulting with the Government on this approach to ensure this opportunity to strengthen protections for young people is clear and contains proportionate obligations that reflect how different digital services operate and the risks they present.

Ongoing risk assessments should occur along with concrete teen safety safeguards, including privacy-protective age estimation, default protective experiences when age is uncertain, under-18 safety policies, safeguards against manipulative or deceptive AI interactions, research-informed crisis response protocols, and accessible parental controls. Together, these measures establish clear and enforceable safety requirements today while ensuring that companies remain responsible for continually assessing risks and strengthening protections as technologies and user behaviors evolve.

To promote accountability and public confidence, legislative frameworks should also require practical independent assessments evaluating how effectively companies have implemented teen safety safeguards and mitigated identified risks. Independent assessments should be underpinned by common standards that support credible, interoperable audits across jurisdictions. Providers should publish a plain-language summary of these assessments on their websites, while appropriately protecting information that could materially increase the risk of misuse or enable circumvention of safety measures. This transparency will help regulators, parents, researchers, and policymakers better understand how companies are addressing teen safety risks in practice and support continuous improvement across the industry.

OpenAI remains committed to working with policymakers, regulators, child safety advocates, parents, educators, researchers, and community organisations in Australia to support legislative frameworks that establish strong protections for teens while supporting responsible innovation.

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 as AI advances, it 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 1 billion people around the world. 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.