

OpenAI

# A European youth safety blueprint for AI chatbots



May 2026

# Foreword

Today's young people will be the first generation to grow up with AI as part of everyday life. It will support how they learn, shape the skills they need for the future, and unlock new ways to create, explore, and express themselves. Getting this right matters. Our responsibility is not only to protect young people from harm, but to ensure they can fully benefit from AI's extraordinary potential.

Access to safe and trustworthy AI should be seen as a right, because AI is a revolutionary technology that will help people unlock their potential and shape their future. Unlike social media — which was fundamentally designed to capture attention and maximize engagement — AI is, at its core, a tool for building, discovery, education, and creativity. That distinction matters. AI can expand opportunity, accelerate learning, and help people shape their futures rather than simply consume content. But precisely because this technology will become so central to economic and civic life, we have a rare window to act earlier, more thoughtfully, and with greater intention than we did with previous technological shifts. This is a unique opportunity to put the right protections, transparency, and safeguards in place from the outset while preserving broad access to AI's benefits.

That is why we believe empowerment and protection must go hand in hand. Young people should be able to engage with AI confidently, safely, and in ways that are appropriate for their age and that support their development and wellbeing. Responsible access requires thoughtful guardrails: protections that reduce risk, safeguards that are proportionate to the technology, and rules that enable trust without limiting the benefits AI can bring.

In Europe, this conversation is taking place within a mature regulatory environment. European and British governments have already established a strong foundation for responsible technology governance through frameworks such as the GDPR and the AI Act. Other laws, including the Digital Services Act and the UK Online Safety Act, are also shaping standards for how online services protect young people. As teenagers embrace AI for learning and creativity, **we have an opportunity to build on this foundation with more targeted and harmonized youth-safety frameworks for AI at the EU level and for the UK without limiting access to beneficial tools.**

Across the region, governments, educators, researchers, and civil society are already helping shape a more thoughtful approach to youth and AI. In Estonia, Greece, Italy and the United Kingdom, we are partnering with governments and

universities to deliver AI tools to students and teachers to accelerate learning and research breakthroughs, while also partnering to develop new tools for understanding AI and learning outcomes. Telefono Azzurro, Italy's leading child safety hotline, has become the first European NGO to partner with OpenAI to deploy AI in support services — enhancing responsiveness while preserving the primacy of human care. These efforts are helping build the evidence base that will inform better policy, better product design, and better protections.

Our commitment in Europe goes beyond deploying responsible products. At OpenAI, we have introduced parental controls to help families guide how ChatGPT works in their homes, updated our Model Spec with Under-18 Principles that detail protections for teens, and have launched an age prediction model — which will come to the EU soon — to determine when an account likely belongs to a user under 18, so that we can apply the right experience and protections. OpenAI is a founding member of the Beneficial AI for Children coalition, launched at the AI Action Summit in Paris in 2025 to advance international collaboration on safeguarding children's development in the age of AI. We joined global and European leaders at the Vatican in signing a declaration to uphold children's rights and dignity in the development and use of AI. And to strengthen the evidence base and to foster a strong ecosystem across the region, we launched a €500,000 grant program to support youth and wellbeing organisations in EMEA.

This European Youth Blueprint builds on those efforts. It reflects our conviction that effective youth protections should be grounded in evidence, harmonized across Europe, tailored to the technology, feasible to implement, and focused on meaningfully improving safety for children and teenagers. It is intended as a contribution to Europe's ongoing regulatory debate: a set of core principles for policymakers considering how to strengthen protections for young people in the age of AI.

Our hope is that this blueprint can serve as a practical foundation for discussion across Europe — helping advance a regulatory approach that is ambitious in protecting young people, clear in its expectations, and capable of supporting innovation that benefits the next generation.

**Ann O'Leary**

Vice President, Global Affairs, OpenAI

**Rafaella Nicolazzi**

EMEA Privacy & Consumers Protection Lead, Global Affairs, OpenAI

# Our blueprint

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

**01** Recognise the positive uses of AI and encourage responsible adoption in education

---

**02** Require AI providers to deliver age-appropriate experiences through proportionate safeguards and privacy-preserving age assurance

---

**03** Mandate AI providers to identify and mitigate risks to minors through under-18 safety policies

---

**04** Stipulate safeguards for protecting minors from manipulative and deceptive AI outputs

---

**05** Establish common standards for accessible parental controls and promote their use

# 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 at home, at school, 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 and Codex, our coding app, are designed for learning and building, not passive scrolling. 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. Europe should avoid one-size-fits-all rules that treat all digital services the same; 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. Early uses can help reduce administrative burdens, support lesson planning, generate materials, and free 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.

Across Europe, educators and institutions are beginning to explore ChatGPT Edu in educational settings, and partnerships with the Estonian government and local schools are helping build evidence on how AI can improve learning outcomes and support student wellbeing. Smart policy should make room for these benefits to grow safely and responsibly.

## European regulation should support:

AI literacy resources for students, parents, teachers, and communities that help young people use AI to support learning rather than replace it; critically assess AI outputs; understand risks and limitations; and use AI transparently and with integrity.

---

Maintain a clear legal distinction between AI tools, including those designed for learning, productivity, and assistance, and other types of digital services where the design, goals, and risks are different.

---

Support for future-ready and equitable pathways to AI fluency, including strong core curricula in national schools, expanded science, technology, engineering, and mathematics, computer science, and career-technical education, open educational resources, modern broadband and devices, and research partnerships that build evidence on what improves learning outcomes and wellbeing.

# Require AI providers to deliver age-appropriate experiences through proportionate safeguards and privacy-preserving age assurance

Regulation should focus on outcomes: ensuring that teenagers receive age-appropriate experiences, while preserving privacy and avoiding unnecessary barriers for all users. A practical framework should rest on two ideas reflecting core European principles such as proportionality, data minimisation, and fundamental rights.

## Identifying teens on AI platforms to treat teens like teens.

AI companies should distinguish between teens and adults on their platforms using privacy-protective, risk-based age prediction tools. These tools should minimise the collection of sensitive personal data while still effectively distinguishing under 18 users. Where possible, these methods might also rely on operating systems, app stores, or other ecosystem-level solutions to determine a user's age. Age prediction helps AI companies ensure that they are applying the right protections to the right users. It facilitates age appropriate experiences and allows AI companies to treat teens like teens. AI companies should provide users with the ability to appeal these decisions if they believe they were made in error.

## Defaulting to protective safeguards when there is doubt about a user's age

AI companies should default to protective safeguards if there is doubt about a user's age. Consistent with our mission to benefit all 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, and in instances where we may not have a good understanding of a user's age, we default to a more protective experience.

### European regulation should support:

A flexible, multi-layered approach to age assurance that enables AI companies to use privacy-protective, risk-based age estimation tools to distinguish teens from adults, minimising the collection of sensitive personal data.

---

A requirement that AI providers apply a more protective default experience whenever they cannot determine with sufficient confidence whether a user is under 18.

# Mandate AI providers to identify and mitigate risks to minors through under-18 safety policies

Teens have specific developmental needs that differ from adults. AI companies should recognise this and not cut corners when it comes to teen safety and wellbeing. AI systems should be designed with age-appropriate protections by default, informed by risk assessments. This means that AI companies should have safety policies for users under 18 that facilitate age-appropriate interactions and evaluate their implementation through testing prior to deployment and through monitoring and enforcement after deployment. Policies should reflect the developmental needs and capabilities of different age groups. Accordingly, the minimum policy baseline for users under 18 should ensure that AI systems:

Do not encourage suicide or self-harm.

---

Prohibit graphic or immersive (e.g., role-playing) sexual and violent content.

---

Do not instruct, encourage, or facilitate (e.g., 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 behaviours through appearance ratings or restrictive diet coaching and age-restricted goods.

---

Do not encourage teens to isolate or keep secrets from trusted adults.

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 policy that describes the company's approach to protecting children and teens, including the safeguards it implements and the parental tools available to families.

## European regulation should consider:

A requirement for AI companies to design their systems with built-in measures to prevent generating content that is unsafe or developmentally inappropriate for kids, including graphic sexual or violent material.

---

A requirement for AI companies to publish child safety policies that explain the safeguards in place and the parental tools available. This transparency builds trust and helps parents make informed choices.

# Stipulate safeguards for protecting minors 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 users' wellbeing rather than exploit immaturity, trust, or dependency.

This means moving beyond traditional safety concepts and prescriptive one-size fits all defaults and instead address the design underpinning AI outputs. For example, regulation should focus on the specific risks policymakers are seeking to prevent — such as emotional dependency to inappropriate relationship dynamics — and require companies to mitigate those risks directly.

AI systems should not be built in ways that intentionally encourage emotional overreliance, compulsive engagement, secrecy from parents or trusted adults, or confusion about whether a user 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 friend, a romantic partner, or a trusted human authority figure.

AI systems used by minors 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.

## European regulation should:

Prohibit manipulative design targeting minors. Regulation should require safeguards that avoid encouraging emotional overreliance, secrecy from parents or trusted adults, or otherwise exploit minors' developmental vulnerabilities.

---

Require age-appropriate AI disclosure. AI systems should make clear that teens are interacting with AI and not humans and include safeguards against deceptive or anthropomorphic design that could mislead children into believing the system is human, conscious, a romantic partner, a personal companion, or a trusted human authority figure.

---

Require relevant wellbeing interventions. AI systems used by minors should include proportionate wellbeing support based on risk and context, such as break reminders, signposting to helplines or crisis resources when distress is identified, and guidance toward offline support from parents, guardians, teachers, or professionals.

# Establish common standards for accessible parental controls and promote their use

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. European regulation should also encourage common industry standards across digital services so parents can navigate parental controls easily and consistently, without facing a different system or learning curve on every platform. 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. 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.

---

*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, as appropriate.

---

*Set blackout hours* to ensure teens take breaks and 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.

Parental tools should be easy to find and actively promoted so parents know they exist and understand how to use them.

## European regulation should embed:

A requirement for AI companies to offer accessible, easy-to-use parental controls — and actively promote them to families — so parents can understand and manage time limits, privacy settings, and feature access.

---

A requirement for notifying by default parents who have their account connected to their teen's account if their teen expresses suicidal intent.

---

A requirement that parents be notified, where appropriate and consistent with the child's age and rights, when key safety or privacy settings are changed.

# Conclusion

We remain committed to ensuring that teen safeguards will be informed by evidence 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 WellBeing 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 protecting them from potential harms.

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