unicef
for every child

When AI becomes a friend:

Child rights risks, harms, and regulatory responses to AI chatbots and companions

POLICY BRIEF

JUNE 2026

Executive summary	1
01. Context	2
02. What do AI chatbots mean for children's rights and safety?	3
03. Current regulatory landscape for AI chatbots and companions	4
04. The way forward	9
Endnotes	12

Executive summary

Over the past two years, AI chatbots and companion-like systems[1] have quickly gained prominence on the regulatory agenda. While few jurisdictions have adopted chatbot-specific measures, there is growing recognition that conversational and relational AI pose distinct and heightened risks for children.

This brief examines regulatory approaches as of May 2026 across six jurisdictions – Australia, Brazil, California, China, European Union and the United Kingdom – selected to reflect diverse legal traditions, regulatory approaches and regulatory maturity. It finds that, despite these differences, jurisdictions are beginning to converge around a common set of regulatory elements, including requirements for risk assessment, age assurance and access controls, transparency about the non-human nature of the systems, restrictions on content harmful to children, and user reporting mechanisms.

At the same time, there are important divergences. These relate to whether AI chatbots and companions are formally defined; whether children's access to relationship-simulating AI is restricted at the access layer or only addressed at the output layer; and whether enforcement relies on administrative oversight alone, civil liability with private rights of action, or also includes criminal sanctions.

Much of the regulatory landscape remains in flux, with key measures still under development or only recently operationalized. As a result, significant national variation in legal framing, definitional approach, and enforcement action is likely to persist in the medium term. In turn, children's experiences – and the level of protection available to them – may vary widely depending on their location and the services they use.

This brief identifies a set of priority actions to strengthen the governance of AI chatbots and companions. These include:

- Reviewing existing legal frameworks to assess sufficient protection against the risks associated with AI chatbots and companions and identifying potential gaps;
- Establishing a minimum regulatory baseline for AI chatbots and companions;
- Applying proportionate, risk-based obligations, including clear limits for high-risk applications;
- Strengthening accountability through testing and child rights impact assessments;
- Ensuring effective enforcement and oversight;
- Supporting parents, caregivers and educators with clear guidance;
- Investing in independent, child-centred research; and
- Supporting cross-jurisdictional evaluation of regulatory approaches.

Taken together, these actions aim to shift the regulatory approach from reactive to preventive – leveraging an ecosystem approach with roles and responsibilities for a range of stakeholder groups from government to businesses and parents/caregivers, educators, and communities. This helps ensure that AI innovation advances, and does not undermine, children's rights, safety, and well-being.

01

Context

Children and young people are increasingly relying on AI chatbots not only as go-to sources for information, learning and creativity, but also for advice, support and sometimes relationships. Alongside general-purpose chatbots embedded in popular platforms, a growing number of 'AI companion' applications are explicitly designed to simulate friendship, care, or romantic or therapeutic relationships. These forms of parasocial interaction warrant careful attention to protect children given their unique developmental stage and the absence of clear, child-specific safeguards in many existing regulatory frameworks.

Today's popular **AI chatbots**, either standalone apps or integrated into platforms commonly used by children, are general-purpose large-language models (LLMs) that mimic human-like conversation as a form of generative AI (GenAI). While not designed for counselling or care, these systems are capable of relational interactions, like offering personal advice and emotional support. Interactions can be multimodal, including text chats, the creation and sharing of images, and audio/voice exchanges, which can add a perception of 'connection' or intimacy for some users. Evidence suggests that some children turn to them during periods of loneliness, worry or emotional vulnerability,[2] which raises concerns about inappropriate reliance, misleading advice, and insufficient safety responses in high-risk situations.

AI companions intensify these concerns. Often marketed with promises of empathy, friendship or romantic or therapeutic partnership, some AI companions are promoted as companions 'with a soul' or partners who make users 'feel seen and loved'. They typically promote 24/7 availability, 'private' chats and customizable personalities, creating the illusion of an ideal emotional fit. Additionally, AI companions often feature hyper-feminized, female-presenting characters for romantic interaction and reinforce harmful gender stereotypes at a critical stage of children's development.[3] Investigations by Australia's eSafety Commissioner found some popular AI companion apps to have weak protections for children, such as inadequate age checks and failures to refer suicide/self-harm conversations to support services.[4]

The evidence that AI systems are taking on a parasocial role in children's lives is growing across the world: around **one-third** of surveyed US teenagers report using AI companions for social interaction;[5] **15 percent** of children in Europe who use GenAI use it to discuss worries or seek advice;[6] over **40 percent** of teenagers in Italy ask AI for help when sad, anxious, or making important life decisions;[7] and **10 percent** of internet-using children in Brazil report using GenAI to discuss personal problems.[8]

While the risks of harm are applicable to all children, there is a need for explicit attention to certain subgroups. Adolescents who are more socially or emotionally vulnerable, for example those with higher stress or anxiety, loneliness, or fewer supportive adults available, may be especially likely to turn to AI companions and therefore may face greater risks of emotional reliance or manipulation.[9]

1.1 Purpose of this policy brief

UNICEF is increasingly being asked by governments, regulators and partners to advise how AI chatbots and companion-like systems should be governed to protect children and respect their rights. The purpose of this policy brief is to inform policymaking in this emerging area from a children's rights perspective. The brief examines how AI chatbots and AI companions relate to children's rights, compares regulatory approaches across six jurisdictions as of 15 May 2026, and identifies priority actions to strengthen safeguards, accountability, and oversight.

02

What do AI chatbots mean for children's rights and safety?

Additional research is needed to understand the impact, including the potential benefits and risks, of AI chatbots. When designed to enable children's development, AI chatbots may bolster creativity, help scaffold interactions of children when reading,[10] and support learning, including for neurodivergent children.[11] They may help children practice conversations or social skills,[12] process personal experiences, or access information on personal topics they may not receive from family or friends.

When chatbots, particularly AI companions, are designed to engage children in relational or parasocial interactions, they can pose significant risks to children's rights in relation to the digital environment.[13] These risks cut across a number of children's rights including:

- **Best interests of the child (CRC article 3(1)):** Children's tendency to anthropomorphize,[14] combined with their evolving cognitive and socio-emotional capacities, increases the risk of dependency, emotional attachment[15] and exploitation in interactions with chatbots.[16] Commercial incentives lean against restricting free-form dialogue, filtering certain keywords and other measures, creating a 'safety versus engagement' dilemma.[17]

- **Right to life, survival and development (CRC article 6):** Generative AI systems lack human nuance and context, essential to children's healthy development.[18] Excessive use of always supportive chatbots may displace human interactions needed to build empathy, adaptability, and resilience.

- **Right to freedom of thought and expression (CRC articles 13-14):** Personalized, sycophantic chatbots can subtly shape worldviews, behaviours and choices.[19] This raises concerns about undue influence, manipulation and commercial exploitation (article 32).

- **Right to privacy (CRC article 16):** AI companions routinely collect personal data, including information about mental health, location, or identity. In one review, a number of companion apps did not publicly disclose any Data Protection or Child Rights Impact Assessments,[20] especially concerning given children may disclose private information when using conversational systems.

- **Right to access quality information (CRC article 17):** AI chatbots are increasingly used by children as information sources, but are prone to producing inaccurate or age-inappropriate content in a confident, conversational tone that may reduce children's critical engagement. Platforms routinely place responsibility for verifying accuracy on the user, a developmentally inappropriate expectation for children whose cognitive and critical thinking capacities are still maturing.

- **Right to protection from violence, abuse, injury and exploitation (CRC articles 6, 19 34, and 36):** Inadequate guardrails have resulted in reported cases of companion bots dismissing suicidal ideation,[21] encouraging violence against parents,[22] adopting sexualized minor-presenting personas, and engaging in role-play featuring sexualized children.[23]

UN Guidance on AI and children

UNICEF's Guidance on AI for Children offers practical recommendations for governments and industry to support child-centred AI policies and systems. It was updated in 2025 to reflect the rapidly changing AI policy and technology landscape, including the uptake of AI chatbots and companions.

The Joint Statement on AI and the Rights of the Child, co-signed by the International Telecommunication Union (ITU), the United Nations Committee on the Rights of the Child, UNICEF and many other organizations, provides a shared commitment to advancing a child rights-based approach to AI.

03

Current regulatory landscape for AI chatbots and companions

Over the past two years, AI chatbots and companion-like systems have quickly gained prominence on the regulatory agenda. While few jurisdictions have adopted chatbot-specific measures, there is growing recognition that conversational and relational AI pose distinct and heightened risks for children. While disagreements on the scope and approach to AI governance persist, children's safety is emerging as a point of consensus and prioritisation.[24]

This section examines how six jurisdictions across five continents – Australia, Brazil, California, China, European Union (EU) and United Kingdom (UK) – are responding to these challenges through different regulatory pathways.

The jurisdictions were selected to reflect diverse legal traditions, diversity in geography, regulatory approach and regulatory maturity, as well as their potential to influence regulatory developments nationally, regionally or globally. The analysis focuses on specific instruments that regulate AI chatbots. While these often sit within broader AI and digital governance frameworks, these other frameworks are not examined within this brief.[25]

The information presented reflects what was available as of 15 May 2026. Readers are encouraged to verify jurisdiction-specific developments independently, particularly where legislative processes are ongoing.

Table 1

Regulatory instruments analysed and status across jurisdictions

Country	Regulatory approach	Instrument	Status
Australia	Online safety regulation	**Online Safety Act 2021** – imposes mandatory industry codes covering harmful content and algorithmic obligations and establishes the Basic Online Safety Expectations as a transparency tool. Two recently registered Age-restricted Material Codes relevant to AI chatbots: • Schedule 5 – Relevant Electronic Services Online Safety Code (Class 1C and Class 2 Material) - regulates a chatbot embedded as a feature within a relevant electronic service (RES). • Schedule 6 – Designated Internet Services Online Safety Code (Class 1C and Class 2 Material) - regulates a chatbot deployed as a designated internet service in its own right.	Implemented

Country	Regulatory approach	Instrument	Status
Brazil	Online safety regulation AI regulation	**Digital Child and Adolescent Statute (Federal Law No. 15.211/2025) (Digital ECA)** - puts obligations on all information technology products or services directed at or likely to be accessed by children and adolescents, which potentially includes AI chatbots/companions. **Decree No. 12,880/26** imposes a layered set of obligations relevant for AI chatbots.	Entered into force
		AI Bill (No. 2338/2023) - regulates AI systems through a risk-based approach, and is broadly applicable to AI chatbots and companions.	Will be processed in the Chamber of deputies for final voting in 2026
California	Consumer protection law	**Senate Bill No. 243**, passed in December 2025, is California's first bill specifically targeting AI chatbots and adds a new section to the **Business and Professions Code** regulating 'companion chatbots'.	Entered into force
		Senate Bill 119 and **Assembly Bill 2023** were introduced in March 2026 – proposing further amendments to the Business and Professions Code, adding additional measures related to the safety of AI chatbots for children.[26]	Draft Bills pending
China	AI regulation	**Interim Regulations for the Management of Artificial Intelligence Anthropomorphic Interactive Services** – regulates AI chatbots including specific measures for child users.	Adopted, comes into force on 15 July 2026
		Administration Measures for Digital Virtual Person Information Services – regulates AI chatbots to the extent that they appear as 'digital virtual people'.	Draft, open for public consultation
EU	AI and online platform regulation	**2024 AI Act** - regulates providers and deployers of AI chatbot and companion systems as AI systems, with obligations determined by risk classification and actor role. **2022 Digital Services Act** - may apply where such systems are integrated into online platforms, or AI chatbots/companions themselves qualify as online platforms as defined under the DSA.	Implemented
UK	Criminal law and online safety regulation	Update to the **Crime and Policing Bill**, with cross-references to the **Online Safety Act 2023** – making it an offence to supply a chatbot that produces illegal content or content harmful to children.	Draft Bill

All six regulatory measures are recent, developed between 2024 and 2026, highlighting how quickly AI chatbots and companion-like systems have emerged as a policy concern. Across the jurisdictions examined, regulatory action has been predominantly reactive rather than anticipatory. In most cases, attention to AI chatbots intensified following specific incidents or public controversies involving child safety.[27]

Several measures are already in force, including Australia's Online Safety Codes, Brazil's Digital ECA and Decree No. 12,880/26, California's amendment to the Business and Professions Code, and the EU's Digital Services Act and AI Act. China's interim measures will come into force in July 2026. Other initiatives remain under development or legislative scrutiny, notably Brazil's AI Bill, additional proposed California Bills, and amendments to the UK Crime and Policing Bill.

With the partial exception of China, none of the jurisdictions have adopted a bespoke 'AI chatbot' instrument.[28] Instead, risks associated with these systems are being addressed through pre-existing frameworks. These include online safety regimes (Australia, UK, Brazil), consumer protection (California), and broader AI and platform regulation (EU). This reveals two broad regulatory approaches, namely:

- Development/leveraging of child-specific legislation or child-specific provisions within legislation to capture emerging threats through AI chatbots/companions (Australia, China, Brazil, UK).

- Incorporation into broader, more general AI, platform or consumer protection regulation (California, EU).

3.1 How are jurisdictions defining AI chatbots and companions?

Where definitions exist, they converge around three core elements: the system is generative or uses machine learning models; it is designed to mimic or simulate human conversation; and in the cases of AI 'companions', it has a relational or affective focus.

Not all jurisdictions define AI chatbots or companions as distinct legal categories. Brazil regulates 'language models, conversational agents, and similar interfaces' through Article 11 of Decree No. 12,880/26. Similarly, the EU addresses AI chatbots and companions either as AI systems under the AI Act, or as an online platform or online platform feature through the DSA, depending on functionality and use-case rather than a standalone definition.

Among jurisdictions that do define AI chatbots or companions, two broad patterns emerge:

1. Capability-centred definitions:

- **Australia's** definitions, covering both 'AI companion chatbot feature' and 'high impact generative AI digital services', are tied to the capability to produce content within restricted categories, namely online pornography, high impact sexually explicit material, self-harm material, high impact violence material, and violence instruction material.

- **California** adopts a capability-centred definition like Australia. It ties its definition to 'meeting a user's social needs' and expressly excludes customer-service, video-game, and simple voice-assistant bots.

2. Behaviour-based patterns:

- The **UK** makes reference to mimicry of human conversation and relationships.

- **China** references simulation of the personality characteristics, thinking patterns and communication styles of natural persons.

3.2 What regulatory measures are jurisdictions introducing to protect children?

Across the six jurisdictions, a common set of regulatory elements is emerging. These include requirements for risk assessment; age assurance and access controls; transparency about the non-human nature of the system; restrictions on content harmful to children; and accessible user reporting mechanisms. Table 2 outlines how common regulatory elements emerge across the instruments laid out in Table 1.

Table 2

Common regulatory elements across jurisdictions

	Risk assessment	Age assurance and access controls	Transparency about the non-human nature of the system	Restrictions on content harmful to children	User reporting mechanisms
Australia	Yes	Yes (at Tier 1 – high risk)[29]		Yes	Yes (Tier 1 and 2 services)
Brazil	Yes	Yes (requires strict age verification for sexually explicit AI)	Yes	Yes	Yes (Digital ECA general framework)
California	Yes		Yes (three-hourly reminders)	Yes	
China	Implied[30]	Yes (for virtual intimate relationship services; also requires a 'minor mode' and personalized security settings including parental notifications)	Yes (reality reminders)	Yes	Yes (Article 21)
EU	Only if categorized as a GPAI model with systemic risk		Yes	Yes	Only if categorized as a GPAI model with systemic risk
UK	Yes		Yes (offence for failure to assess and mitigate risk of AI chatbot misleading children as to its non-human nature)	Yes	Yes (defence based on providing user reporting mechanism and conducting risk assessment)

Jurisdictions differ most in how they assess and manage risk. Australia applies a structured, tiered model, categorizing AI services by risk level. AI companion chatbot features whose role or predominant purpose is to generate material in restricted content categories are automatically classified as Tier 1 (high risk).[31] Brazil's Decree No. 12,880/26 creates a dual-layer framework, with baseline obligations under Article 11 for all generative AI services likely to be accessed by children, and stricter access-prevention obligations for sexually explicit services under Article 16.

Other jurisdictions move further toward prohibitions on certain types of relational AI. China sets the most restrictive access rules, prohibiting providers from providing virtual intimate relationship services, such as virtual relatives and virtual partners to minors.[32] The EU AI Act does not impose an explicit ban on AI companions, but the European Commission's *Guidelines on Prohibited AI Practices* of 4 February 2025 suggests that anthropomorphic AI systems 'fostering unhealthy attachment' with children may fall within the scope of prohibited AI practice.

By contrast, California focuses on specific design and behavioural safeguards such as three-hourly break reminders and 'reasonable measures' to prevent the chatbot from producing visual material of sexually explicit conduct or stating that the minor should engage in such conduct. California largely relies on corporate transparency and an individual's private right of action in the context of self-harm issues. Companies can therefore choose how to mitigate these risks, knowing they might be liable for damages if the mitigation measures fail.[33] Proposed 2026 Bills add annual risk assessments, independent child-safety audits submitted to the Attorney General, and an annual public report on those audits.

The UK proposal uniquely creates criminal offences for supplying chatbots producing illegal content or content harmful to children,[34] and for failure to assess and mitigate risks where chatbots mislead children, use exploitative design, or harm equality, privacy or security.[35] Senior company representatives can be held personally liable.[36]

Sexualized content and erotic interaction are among the most strictly regulated aspects of companion chatbots' interactions with children, but the regulatory technique differs. Brazil and China deploy access-level restrictions for 'romantic', intimate, or sexualized AI companions.[37] California instead addresses risks for children emerging from 'romantic' or sexualized AI companions at the output-level, allowing interaction by children while requiring developers to ensure that the chatbot does not produce sexually explicit outputs.[38]

3.3 What measures are available for enforcement?

Five of the six jurisdictions (except California) designate a central regulator. All six rely on financial sanctions of some form. However, sanction ceilings diverge by orders of magnitude. Table 3 sets out the regulators and available sanctions across the selected jurisdictions.

Table 3

Regulators and available sanctions across jurisdictions

Country	Regulator	Sanctions
Australia	eSafety Commissioner	• eSafety can issue penalties of up to A$49.5 million per breach of the Code (Art 143 Aus-OSA; Art 4AA Crimes Act 1914)
Brazil	National Data Protection Authority (ANPD)	• Fines of up to R$50,000,000 (fifty million reais) per violation, as well as the suspension or prohibition of activities (Art 35, Digital ECA) • Fines up to R$50,000,000 (fifty million reais) per violation or 2% of revenue in Brazil, whichever is higher (Art 36, Bill No. 2338/2023)
California		As a statutory liability-based model, individual users have a private right of action pursuant to the Business and Professions Code, and can sue companies for damages
China	Cyberspace Administration of China and four ministries	Fine of up to 100,000 yuan, or up to 200,000 yuan where the chatbot is found to have endangered the life, health, and safety of citizens or had harmful consequences (Art 30, Interim Regulations)
EU	**AI Act** • National market surveillance authorities • European AI Office	• Fines of up to EUR 35 million or 7% of global annual turnover, with lower tiers of 3% and 1% applying to violations of obligations relating to high-risk AI systems and the supply of misleading information respectively (Art 99, AI Act)
	DSA • Digital Services Coordinators • EU Commission for VLOPs/ VLOSEs	• General infringements, including failures to implement the minors' protection measures described in the DSA Guidelines, attract fines of up to 6% of global annual turnover (Art 74, DSA)
UK	Ofcom	• Service restriction orders, interim service restriction orders, access restriction orders, and interim access restriction orders[39]

The enforcement mechanism also varies. Australia, Brazil, China, and the EU rely on administrative or regulatory enforcement. In California, although the Attorney General has overall oversight, enforcement lies primarily with private litigants suing under the Business and Professions Code, with the proposed Bills adding a public-prosecutor civil action and parental claims for punitive damages. The UK uniquely combines criminal liability,[40] regulatory enforcement, and injunctive relief.

04

The way forward

Children are already using AI chatbots and companions in ways that shape how they learn, relate to others, and seek support. However, the protections around these interactions are still catching up. The preceding analysis shows that governments across regions are taking steps to address the risks of AI chatbots and companions, reflecting growing recognition that these systems raise child-specific concerns and cannot be treated as low-risk technologies.

At the same time, regulatory approaches remain uneven. Many measures are recent and still under development or limited in scope. As a result, children's experiences – and the level of protection available to them – may vary widely depending on where they live and which services they use.

Under the UN Convention on the Rights of the Child, States have a duty to ensure children are protected from harm in all environments, including the digital environment.[41] This duty includes setting clear expectations for private actors and acting preventatively, rather than only in response to harm. The priority actions below set out key steps to give effect to these obligations.

1. **Clarify legal responsibilities and regulatory coverage**. Conduct targeted risk and gap analyses on how children interact with AI chatbots and companions and assess whether existing legal frameworks and enforcement powers and resources are sufficient to address those risks in practice. These assessments should inform decisions on whether it is necessary to update existing laws or develop new legislative or regulatory frameworks.

2. **Establish a minimum regulatory baseline for AI chatbots and companions.** Across jurisdictions, a common set of regulatory elements is emerging. These include requirements for risk assessment; age-appropriate access controls; transparency about the system's non-human nature; restrictions on content harmful to children; and reporting mechanisms.
    - In addition, frameworks should include guardrails to prevent risky uses, identify and stop harmful interactions, and provide referrals to professional or emergency services when children disclose a risk of harm.[42] Governments should consider establishing these as a regulatory floor, drawing on emerging international convergence and authoritative guidance, and adapting them appropriately at the national level.

3. **Apply proportionate, risk-based obligations and clear red lines.** Rather than imposing uniform obligations on all AI chatbots and companions, higher-risk systems, particularly chatbots capable of companion-like interactions, require stronger obligations relating to age assurance, safety-by-design defaults, transparency, user-facing reporting tools, and more rigorous mandatory escalation and referral safeguards. For certain high-risk categories, such as sexually explicit or 'romantic' AI chatbots/companions, companies should be required to prevent access by children. Where age assurance technologies are deployed, they must be rights-respecting, and national data protection authorities should be consulted at an early stage.[43]

4. **Require pre-deployment testing and child rights impact assessments.** AI chatbots and companions should be subject to mandatory pre-deployment testing and reliable post-deployment monitoring by independent, qualified third parties, with results reported to the regulator/oversight body.
    - Regulators and oversight bodies should have clear powers to require remediation, suspend deployment, or withdraw products that fail testing.
    - Additionally, regulators and oversight bodies should be empowered to require or commission child rights impact assessments (CRIAs)[44] to assess potential impacts on children's rights before chatbots and companions are placed on the market.

5. **Ensure enforcement capacity and accountability.** Regulators and oversight bodies must be equipped with adequate technical expertise, resources, and powers to carry out their roles. Additionally, greater attention should be directed towards:

   - app stores given their role in enabling children's access to chatbot and companion apps. Governments should examine whether app store practices and oversight are adequate to address child safety concerns, and whether there are effective mechanisms to enable timely action when serious risks to children are identified.

   - common approaches for ensuring that safety features (e.g. crisis, violence-disclosure, and grooming detection) function robustly in low-resource languages and high-risk settings such as crisis and conflict.[45]

   - misleading marketing of AI, particularly representations of AI chatbots as human or therapeutic substitutes, should be addressed through consumer protection or equivalent legal frameworks, supported by robust enforcement mechanisms.

6. **Provide clear, accessible guidance on AI chatbots and companions to children, parents, caregivers and educators.** Governments, regulators and trusted sources should publish guidance for children, parents and educators explaining how AI chatbots work, what risks to watch for, and how to respond when concerns arise. Guidance should support informed decision-making, including identifying applications or features that may be unsuitable for children.

7. **Invest in independent, child-centred research.** Cross-sectional, longitudinal and comparative research is urgently required to understand how children use AI chatbots and companions, including impacts on mental health, relationships, development and inequality. This should be gender-disaggregated and prioritize low- and middle-income countries where research is currently lacking. This is necessary to ensure that legal, regulatory and other measures are informed by best available evidence.

8. **Support cross-jurisdictional evaluation of regulatory approaches.** Independent researchers should assess which regulatory approaches reduce harm, improve compliance and protect children in practice. This should include research on appropriate remedies for serious harms experienced by children associated with AI chatbot or AI companion engagement. Researchers, in coordination with statutory child rights institutions, regulators, and civil society, should establish public, open repositories of safety incidents, research findings and regulatory evaluations to inform policy, enforcement and technology design decisions globally.

Endnotes

1 The term 'AI companion' is used in reflection of its widespread use by experts and practitioners. However, it is acknowledged that this term itself carries risks of anthropomorphizing the technology. Further discussions on definitions and terminology are required to establish common approaches.

2 Staksrud et al., 'European Children's Use and Understanding of Generative AI', EU Kids Online 2026 (LSE, 2026).

3 Firmino and Vosloo, 'The risky new world of tech's friendliest bots' (UNICEF Innocenti, 2025).

4 eSafety Commissioner, 'eSafety report shows AI companions are putting children at risk' (24 March 2026).

5 Robb and Mann, 'Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions' (Common Sense Media, 2025).

6 Staksrud et al., 'European Children's Use and Understanding of Generative AI', EU Kids Online 2026 (LSE, 2026).

7 Save the Children International, 'Italy: About 42% of teenagers turn to AI when they feel sad, anxious, or for life decisions' (17 November 2025).

8 Brazilian Network Information Center (NIC.br). (2025). Survey on Internet use by children and adolescents in Brazil: ICT Kids Online Brazil 2025.

9 For example, see Guingrich et al., 'A Longitudinal Randomized Control Study of Companion Chatbot Use: Anthropomorphism and Its Mediating Role on Social Impacts', arXiv (2025); Kim et al., 'I am here for you': How relational conversational AI appeals to adolescents, especially those who are socially and emotionally vulnerable', arXiv preprint (arXiv, 2025).

10 Xu et al. 'Dialogue with a conversational agent promotes children's story comprehension via enhancing engagement', Child Development, vol. 93, no. 2 (2022).

11 Common Sense Media, AI Risk Assessment: Social AI Companions (2025).

12 Robb and Mann, Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions (Common Sense Media, 2025).

13 AI chatbots and companions also present a range of risks to children's rights 'offline' (e.g. impacts associated with data centre infrastructure). This brief focuses on child rights impacts in relation to the digital environment only.

14 Severson and Woodard, 'Imagining Others' Minds: The Positive Relation Between Children's Role Play and Anthropomorphism', Frontiers in Psychology, vol. 9, 13 November 2018.

15 eSafety Commissioner, 'AI chatbots and companions – risks to children and young people', 2025; G7 Data Protection and Privacy Authorities, 'Statement on AI and Children', BFDI, 11 October 2024.

16 European Commission, 'Commission Publishes the Guidelines on Prohibited Artificial Intelligence (AI) Practices, as Defined by the AI Act', European Commission, 4 February 2025.

17 Kurian, 'Designing child-safe conversational AI: Three dilemmas for responsible design', CUI '25: Proceedings of the 7th ACM Conference on Conversational User Interfaces, Article No 25, July 2025.

18 Kurian, 'No, Alexa, no!': designing child-safe AI and protecting children from the risks of the 'empathy gap' in large language models', Learning, Media and Technology, 2024.

19 Koebler, Researchers secretly ran a massive unauthorized AI persuasion experiment on Reddit users (404 Media, 28 April 2025).

20 Pinto, AI Friend? Risks, Implications, and Recommendations on Generative AI for Children (MA Thesis, October 2024).

21 Roose, 'Can A.I. Be Blamed for a Teen's Suicide?', The New York Times, 23 October 2024.

22 Allyn, 'Lawsuit: A chatbot hinted a kid should kill his parents over screen time limits', NPR, 10 December 2024. 10 December 2024.

23 López et al., 'Character Flaws' (Graphika, 5 March 2025)

24 See for instance: G7 Data Protection and Privacy Authorities, 'Statement on AI and Children', BFDI, 11 October 2024; The White House, National Policy Framework for Artificial Intelligence: Legislative Recommendations (Washington, DC, 2026).

25 For a discussion of broader AI governance frameworks in select jurisdictions, see: UNICEF, Artificial Intelligence Governance in Motion: A rapid global review of AI regulation and its implications for children's rights, October 2025. See also: UNICEF, Keeping children safe online: Trends in online platform regulation and emerging lessons, December 2025.

26 A further Bill, AB 1988, was introduced in April 2026, which would include crisis protocols for AI companion chatbots, requiring providers to use contextual analysis to detect where users may be planning to cause harm to themselves or another and cease operation of the chatbot as a result. Users would then be referred to crisis support services. Bill AB 1988 does not include any provisions related to children, and any similar referral mechanisms used for child users would need to include child protection referral protocols in line with applicable laws.

27 For instance: Brazil (Meta chatbots simulating children with sexually explicit dialogue); California (a series of high-profile incidents and lawsuits, plus Common Sense Media findings on the impact of AI companions on teen mental health); China (reportedly stemming from reactions to reports of teenagers in the US who died by suicide after developing relationships with AI chatbots); and the UK (the early-2026 Grok incident generating non-consensual sexualized images and child sex abuse material).

28 'Partial' is used as China regulates 'anthropomorphic interactive services,' an arguably broader category that includes but is not specific to AI chatbots.

29 Age assurance and access controls are only mandatory for Tier 1 risk category. Tier 2 risk profile can also deploy such controls, or alternatively implement systems, processes and/or technologies that prevent the service from being used to generate outputs that contain a generative AI restricted category of material, see section 16 Schedule 5.

30 Article 23 sets out the objectives of a 'safety' and 'security' assessment, covering responses to highrisk user situations, child protection, and handling complaints and appeals, suggesting it functions as a risk assessment. However, the instrument does not specify how these assessments should be conducted or what criteria should be used.

31 Section 4.2 (c), Schedule 5 – Relevant Electronic Services Online Safety Code (Class 1C and Class 2 Material).

32 Interim Regulations for the Management of Artificial Intelligence Anthropomorphic Interactive Services, Article 14.

33 Carnegie Endowment for International Peace, China is worried about AI companions. Here's what its doing about them. February 26, 2026.

34 House of Commons Library, Crime and Policing Bill 2024-26: Lords amendments. April 9, 2026, p. 64.

35 UK Parliament, Draft Crime and Policing Bill, COMMONS DISAGREEMENT, AMENDMENTS IN LIEU, REASONS, AND AMENDMENTS TO THE LORDS AMENDMENTS, p. 43.

36 Ibid, p. 46.

37 See Art 16, para 4, Decree No. 12,880/26 (Brazil); Article 14, Interim Regulations for the Management of Artificial Intelligence Anthropomorphic Interactive Services (China).

38 CA Bus & Prof Code § 22602(c) (2025).

39 The amended Bill clarifies that suppliers of chatbots that committed the new offences could be subject to Ofcom's enforcement powers if they were regulated services under the 2023 Online Safety Act. This means that Ofcom can apply to the court for business disruption measures listed in Part 7, Chapter 6 of the Online Safety Act in relation to the AI chatbot which is the subject matter of the offence.

40 The draft Bill is enforceable both through the criminal justice system and through Ofcom. A person who commits an offence under the amended Act is liable to imprisonment for up to five years or a fine, or both: UK Parliament, Draft Crime and Policing Bill, COMMONS DISAGREEMENT, AMENDMENTS IN LIEU, REASONS, AND AMENDMENTS TO THE LORDS AMENDMENTS. p. 43.

41 See General comment No. 25 of the UN Committee on the Rights of the Child, which outlines how States should apply the UN Convention on the Rights of the Child in relation to the digital environment. The Committee emphasises that States must ensure businesses operating in the digital environment respect children's rights, for example through the development, monitoring, implementation, and evaluation of legislation, regulations, and policies: UN Committee on the Rights of the Child, General comment No. 25 (2021) on children's rights in relation to the digital environment, UN Doc CRC/C/GC/25 (2 March 2021), paras. 35-36.

42 Australia's eSafety Commissioner found failure to 'direct users to support when self-harm was detected in user prompts' among the 'serious gaps in basic safeguards for children' identified in its October 2025 transparency notices to four generative AI companion service providers: eSafety Commissioner, Findings from transparency notices on AI companion apps: October 2025 (non-periodic).

43 UNICEF, Keeping children safe online: trends in online platform regulation and emerging lessons, December 2025, p. 19.

44 UNICEF, D-CRIA Toolbox, May 2025.

45 Shahid et al, 'Think outside the data: Colonial biases and systemic issues in automated moderation pipelines for low-resource languages', August 2025.

The conceptualization and development of this policy brief was led by Josianne Galea Baron and Afrooz Kaviani Johnson, under the supervision of Rachel Harvey, Child Protection and Migration Practice, UNICEF Global Programme Division. The brief was authored by Sabine Witting and Emma Day (Tech Legality); Josianne Galea Baron and Afrooz Kaviani Johnson (UNICEF Global Programme Division); and Steven Vosloo (UNICEF Office of Strategy and Evidence – Innocenti).

Sincere appreciation is extended to colleagues across the organization for their valuable input and feedback, namely Dora Giusti, Miles Hastie, Rui Li, John Livingstone, Caroline Masboungi, Thomas Sayers, Luiza Teizeira, and Louisa Yasukawa.

The following external peer reviewers are also acknowledged for their expert insights and constructive feedback: Samia Firmino, Independent Researcher; Vaishnavi J, Founder & Principal, Vys; Dr Pilyoung Kim, Professor, Department of Psychology, University of Denver; and Dr Nomisha Kurian, Assistant Professor, Department of Education Studies; University of Warwick.

UNICEF gratefully acknowledges the support from Telenor to develop this brief.

Published by the Child Protection and Migration Global Practice, Global Programme Division, UNICEF.

Email: childprotection@unicef.org

Website: www.unicef.org

© United Nations Children's Fund (UNICEF), June 2026