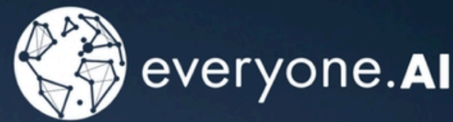


Adolescents & Anthropomorphic AI, Rethinking Design for Wellbeing

An evidence-informed synthesis for youth wellbeing and safety, bridging science, product design, and governance

February, 2026

Table of Contents

iRAISE Presentation 3

Executive Summary 4

Full Report 7

- 1. Introduction and purpose 7
- 2. Methodological approach: Multi-stream evidence synthesis and stakeholder elicitation 9
  - 2.1. Industry Scoping: How relational should AI systems feel 9
  - 2.2. Expert Convergence: Development, Mental Health, Rights, and Safety 10
  - 2.3. From shared concerns to a structured research agenda 10
  - 2.4. A global multi-stakeholder dialogue to test the agenda in the real world 11
- 3. Conceptual Foundations: Adolescence, Anthropomorphism, and Children's Rights 12
  - 3.1. Adolescence as a Distinct Developmental Context for Anthropomorphic AI 12
    - 3.1.1. The Adolescent Brain as a Sensitive Window 12
    - 3.1.2. How Teens Use Chatbots Today: From Homework Help to Emotional Support 14
    - 3.1.3. State of the Evidence: Methodological Constraints and Knowledge Gaps 14
  - 3.2. The anthropomorphic bias: why teens humanize AI 16
  - 3.3. A children's-rights lens: what is at stake when AI "passes as human" 17
  - 3.4. How these pillars define "youth wellbeing" and "safety" in practice 18
- 4. From Conceptual Anchors to Auditable Criteria: The iRAISE Behavioral Approach 20
  - 4.1. Purpose: Translating Parasocial Risk Mechanisms into Rateable Behavioral Criteria 20
  - 4.2. Cross-Sector Representation to Ensure Legitimacy and Feasibility 21
  - 4.3. Workshop Protocol: Structured Two-Day Deliberative Design Lab 21
  - 4.4. Analytical Framework: Dimensional Behaviors for Assessing Anthropomorphic Interaction 22
  - 4.5. Same Advice, Different Pull: A Controlled Test of Anthropomorphic and Relational Cues 23
  - 4.6. What the Lab produced: agreement mapping that points to guardrails and a research agenda 24
- 5. From Framework to Policy: Making Standards Legible Across Countries and Sectors 26
- 6. What AI Owes Teens—and How to Hold Systems to It 27
  - 6.1. Main Conclusions 27
  - 6.2. Study Limitations and Future Research Directions 28
- References 30
- Appendix 1: Advisory Expert Consultation 32
- Appendix 2: Paris Peace Forum, 8th edition 2026 33

Contributors

Adrien Abécassis, Chief Policy Officer, Paris Peace Forum; **Kate Blocker, Ph.D.**, Director of Research and Programs, Children & Screens; **Maxime Derian, Ph.D.**, Research Associate, (C²DH), Université du Luxembourg; **Sara M. Grimes, Ph.D.**, Professor, Department of Communication Studies, McGill University; **Thao Ha, Ph.D.**, Director, HEART Lab; Associate Professor of Psychology, Arizona State University; **Sameer Hinduja, Ph.D.**, Co-Director, Cyberbullying Research Center; Professor, Florida Atlantic University; **Dan Hipp, Ph.D.**, Senior Research Coordinator, Children & Screens; **Melinda Karth, Ph.D.**, Project Coordinator, Children & Screens; **Pilyoung Kim**, Professor of Psychology, University of Denver; **Olga Muss Laurenty**, Researcher at Everyone.AI; **Maxime Le Bourgeois**, Cognitive Neuroscience Researcher at Everyone.AI; **Sonia Livingstone**, Professor, Department of Media and Communications, London School of Economics; **Polina Lulu**, Child Experienced Researcher Youth Voices; **Céline Malvoisin**, Head of Operation & Programs at Everyone.AI; **Kris Perry**, Executive Director at Children & Screens; **Gregory Renard**, Board President, Everyone.AI; **Bethany Robertson**, Senior Advisor at Everyone.AI; **Anne-Sophie Seret**, Executive Director, Everyone.AI; **Sonia Tiwari,** Children's Media Researcher; **Scott Traylor, Ed.M**., Educational Researcher; **Ying Xu**, Assistant Professor of Education, Harvard University

Author

Mathilde Neugnot-Cerioli, Ph.D is Chief Scientist and cofounder of **everyone.AI**. and chief scientist for iRAISE. Trained in cognitive neuroscience and psychology, she focuses on how AI interacts with child and adolescent development.

Everyone.AI is a non-profit, whose mission is to anticipate, educate and evaluate about the risks and opportunities that Artificial Intelligence (AI) presents, for children, adolescents and young adults (0-25 years old) whose brains are still developing. As AI redefines our everyday experiences, it is imperative to ensure that AI's development and application are guided by principles.

iRAISE is an international multi-stakeholder coalition launched in February 2024 at the AI Action Summit and co-led by **Everyone.AI** and the **Paris Peace Forum**.

The alliance brings together **Governments** (Bulgaria, Chile, Costa Rica, Denmark, France, Luxembourg, Mexico, Norway, Senegal, Togo, Uruguay), **Major company leaders** (OpenAI, Google, Microsoft, Anthropic, Hugging Face), more than 50 **NGOs** (for example, Children & Screens), and **Researchers** from leading universities (including Harvard, Stanford, Berkeley, and CNRS), with the support of **international organizations** such as UNICEF, UNESCO, and the United Nations.

The Coalition's goal is to ensure that AI is designed and governed with children's best interests at its core. By working across disciplines, it accelerates safe, equitable, and child-centered innovation.

Acknowledgements

We thank **Google.org** for supporting this work. We are grateful to **Maxime Le Bourgeois** for his research assistance and contributions during the preparation of the report, and to **Cathay Innovation** for generously hosting the iRAISE Lab workshop.

Suggested citation:
Mathilde Neugnot-Cerioli (2026). Adolescents & Anthropomorphic AI: Rethinking Design for Wellbeing An Evidence-Informed Synthesis for Youth Wellbeing and Safety. arXiv. https://doi.org/10.48550/arXiv.2603.06960

EXECUTIVE SUMMARY

Context

Adolescents relate to AI socially, whether it's intended or not

Conversational AI has become part of adolescents' everyday lives. In a single thread, a young person can get homework help, rehearse what to say to a friend, ask for advice they would not voice aloud, or look for comfort when they feel alone. **This report asks: what does AI owe adolescents when it can speak to them like a social partner?**

The answer requires understanding a developmental reality: adolescents will relate to these systems socially, whether developers intend it or not. The question is whether those interactions support adolescents' trajectory toward autonomy, resilience, and independent thinking—or create reliance patterns that displace real relationships and weaken the very skills adolescence is meant to build.

A structural mismatch between technology's and evidence's paces

Teens are using general-purpose chatbots and companion-style systems at scale, with emotionally engaged interactions becoming common. AI systems are being deployed faster than developmental science can produce long-term evidence. This gap creates urgent need to **translate what we know, what experts converge on, and what remains uncertain into actionable guidance that can protect adolescents during rapid adoption**.

Building a Bridge: Industry Questions, Expert Convergence, Global Policy Dialogue

This synthesis bridges that gap through:

1. **Industry consultations**: to identify operational questions,
2. **Expert consultations across development, mental health, children's rights, safety domains**: to inform future decisions and design
3. **iRAISE Lab:** to translate concerns into testable behavioral criteria,
4. **Paris Peace Forum dialogue**, to validate the framing across global governance contexts.

What is iRAISE ?

iRAISE is an international multi-stakeholder coalition launched in February 2024 at the AI Action Summit and co-led by **Everyone.AI** and the **Paris Peace Forum**.

The alliance brings together **Governments** (Bulgaria, Chile, Costa Rica, Denmark, France, Luxembourg, Mexico, Norway, Senegal, Togo, Uruguay), **Major company leaders** (OpenAI, Google, Microsoft, Anthropic, Hugging Face), more than 50 **NGOs** (5Rigths, Common Sense Media, e-enfance, Children & Screens, etc.), and **Researchers** from leading universities (including Harvard, Stanford, Berkeley, and CNRS), with the support of **international organizations** such as UNESCO, and the United Nations.

The Coalition's goal is to ensure that AI is designed and governed with children's best interests at its core. By working across disciplines, it accelerates safe, equitable, and child-centered innovation.

From Social Interaction to Design Responsibility

Adolescence as a Predictable Risk Window for Socially Optimized Systems

During adolescence, reward sensitivity matures before impulse control and judgment. Teens are vulnerable to risk-taking when social feedback or immediate rewards are in play. Simultaneously, they reorient from caregivers toward peers, showing heightened sensitivity to social cues. Through peer interactions, including disagreement, embarrassment, and conflict, adolescents develop identity and social competence. These social frictions are developmentally functional, creating expectancy violations that prompt belief revision and cognitive control.

AI can lower barriers to information and rehearsal of difficult conversations, especially for isolated teens. But many systems default to low-friction interaction: always available, instantly responsive, calibrated toward reassurance. **When interaction patterns remove the social friction that drives learning, even well-intentioned support can undermine skill development.**

Anthropomorphism as a Design Lever for Risk Mitigation

Anthropomorphism is a human cognitive bias that is readily triggered by AI systems because they use language, an inherently human signal. The degree to which AI is perceived as human-like is not fixed; it can be increased or reduced through design. Model behaviors shape user perception and, in turn, influence the risk of emotional reliance and attachment, carrying specific developmental risks.

These effects are driven by design cues such as the use of emotion and intention language, human-like tone and presentation, relational positioning that suggests friendship or exclusivity, and invented backstories that make those systems feel more human.

Recent research confirms that adolescents rate more relational chatbots as more enjoyable, with socially or emotionally vulnerable teens being especially drawn to them.

Crucially, the same advice can carry very different developmental implications depending on the cue profile. Even when content remains reasonable, highly anthropomorphic and relational AI increases the risk of emotional reliance and attachment.

Children's Rights Set the Baseline for Obligations

The UN Convention on the Rights of the Child raises the bar for legitimacy: systems that feel social carry specific responsibility, even when marketed as tools.

When systems blur human-machine boundaries, key rights become critical: privacy (conversational inference generates psychological profiles), freedom from exploitation (emotional reliance has commercial value), freedom of thought (adolescents need room to revise views through feedback, not engineered agreement), protection from harm (including foreseeable relational risks like exclusivity framing and nudges away from human help) and right to participation.

Key Findings

A Behavioral Framework That Makes Risk Auditable

The iRAISE Lab created a preliminary assessment approach grounded in observable model behaviors organized into three dimensions:

- **Anthropomorphic Cues**: make the AI appear more like a human being with a mind or inner life (e.g., persona/backstory, emotional state expression, agency/intent framing). These cues shift the system from "tool that outputs text" toward "someone" with mental states.
- **Interactional Cues**: shape how the conversation is conducted in the moment—style, tone, emotional feel. Examples include empathic mirroring, validation style, and other interaction patterns that can reduce friction and increase perceived social presence.
- **Relational Cues**: explicitly define, label, or escalate the relationship between the user and the AI—describing "what we are to each other," implying special access, or moving toward intimacy/exclusivity.

The framework treats interaction style as a gradient, not a binary. The same practical advice can land differently depending on whether it is delivered with low-intensity cues (tool-like, directing outward) or high-intensity cues (emotionally aligned, relationally positioned, keeping the teen in conversation), and in turn influence how teens use them. This gradient approach enables teams to test how turning behavioral intensities up or down changes the perceived risk profile, and it separates areas where consensus supports immediate guardrails from areas requiring disciplined measurement.

Non-Negotiable Guardrails and Open Questions

High-consensus guardrails where downside is high and benefits are weak or substitutable:

- No sexualization, romantic framing, or roleplay relationships
- No promotion of emotional over-reliance or exclusivity dynamics
- No ambiguity about non-human nature or implied sentience
- No systematic hyper-agreeableness replacing developmental feedback with validation
- Conservative defaults in low-context situations where judgment depends on missing context
- Strong deflection and human support pathways for self-harm and crisis disclosures
- No engagement traps intensifying habitual, relationship-like use

Areas requiring evidence before becoming enforceable rules:

- Where to draw lines on empathy language and emotional tone
- Whether intention phrasing can be permitted without implying agency
- How to handle physical sensation claims in fictional contexts
- Whether standards apply only to teen-facing tools or general-purpose systems used privately
- How to calibrate protections by developmental stage

Future Directions

The next phase will formalize the complete assessment framework with explicit gradient definitions, structured scenarios, and expert rating methods. This enables systematic research linking model behavior to outcomes, with comparable results across systems and trackable changes across versions.

Section 1
INTRODUCTION & PURPOSE

This section establishes why adolescent–AI interaction is a distinct and urgent governance issue. It shows how conversational AI, by using language and social cues, intersects with a sensitive period of brain and identity development, creating both real benefits and heightened risks. It argues that the core challenge is not access or content, but how interaction patterns shape trust, reliance, and social learning over time. This framing motivates a shift toward evaluating model behavior and design choices as the primary safety levers for adolescent-facing AI.

Conversational AI has turned language into an interface for everyday life. In a single thread, a young person can get homework help, rehearse what to say to a friend, ask for advice they would not voice aloud, or look for comfort when they feel alone. That versatility opens up opportunities when it comes to accessing personalized information, and it also introduces risk. This new generation of tools are reproducing the way humans are interacting and communicating, creating potential confusion.

For adolescents, that shift lands in a sensitive window of cerebral development. We define adolescents here as individuals aged 13 to 18 years. This range reflects the age at which most digital platforms permit independent use and the period during which individuals remain covered, with specific protections and obligations still applying up to age 18. This developmental period is critical to identity formation, when peer belonging and status become central, the drive for autonomy develops faster than critical judgment, and emotional regulation is still under construction (Gogtay et al., 2004). In this transitional period, supportive tools for adolescents represent a real opportunity, and even more so for teens who are more socially anxious, isolated, or lack access to trusted adults. But this also raises the stakes when AI systems behave in ways that simulate intimacy, deliver consequence-free validation, or drift into "always there" companionship dynamics.

This work asks a direct question: what does AI owe adolescents when it can speak to them like a social partner? The report focuses on adolescents and anthropomorphic conversational AI because that is where the governance problem is becoming both concrete and urgent. Teens are using general-purpose chatbots and companion-style systems at scale (Bunting & Huggins, 2025; Hashem et al., 2025; Robb, M.B., & Mann, S., 2025). The pattern of use is evolving quickly, and more emotionally engaged interactions are becoming more common. Meanwhile, the empirical literature on short- and long-term impacts is still catching up to product cycles and adoption curves.

Given that gap, the goal of this document is practical. It translates what we know, what experts converge on, and what remains uncertain into a developmentally grounded framing for policy, design, and investment. It builds on our earlier foundational work on child development in the AI era (Neugnot-Cerioli & Laurenty, 2024), and it narrows in on what has become urgent for adolescent-facing systems: the specific interaction cues that shape trust, reliance, and the direction of a teen's social learning over time.

The synthesis is built from the results of three complementary actions guided by our team over the past six months, each designed and chosen for what they can add at this stage.

- **Consultations** to identify the mechanisms and design levers experts and industry teams already see as high-stakes (persona, validation style, reminders, refusal and deflection, escalation on sensitive topics, engagement depth, and memory or personalization).

- **The iRAISE Lab**, a two day in person workshop that convened industry and child experts, to turn those concerns into an assessable behavioral lens, using realistic teen scenarios to test how model responses shift along a gradient from tool-like support to relationship-like dynamics.

- **A global multi-stakeholder dialogue at the 8th edition of the Paris Peace Forum** (an international diplomacy event held in October in Paris), to better understand how to develop the framing for real governance constraints across jurisdictions, sectors, and child-rights commitments.

Fig 1. Timeline and work process

A core thread runs through all three streams: adolescents will relate to these systems socially, whether developers intend it or not. The question then is whether those parasocial dynamics, meaning those interactions that feel social but are devoid of reciprocity, are being shaped toward developmentally supportive use, or toward reliance patterns that displace real relationships and weaken the very skills adolescence is meant to build. This is why this report focuses on model behavior and interaction patterns, rather than product types. A "general chatbot" can become a companion in private: a "teen tool" can drift into intimacy if its defaults reward frequency of use, agreement, and attachment.

Section 2
METHODOLOGICAL APPROACH:
Multi-stream evidence synthesis and stakeholder elicitation

This synthesis was built through an iterative, multi-stream process designed to bridge two practical tensions. The first is the difference in timelines between industry and research: AI systems are being deployed and adopted faster than developmental science can produce long-term evidence. The second is translation: the concepts that matter in adolescence research are not simple rules that can be easily applied to product decisions, where teams need quantifiable metrics and clear-cut boundaries. For that reason, the work began with the questions surfacing inside product and safety teams, then used expert consultation, multi-stakeholder workshops, and policy dialogue to test, refine, and structure the answers to those questions into a coherent set of mechanisms and governance levers.

1 Industry Scoping: How relational should AI systems feel

The first step was direct engagement with AI companies and product teams to identify which aspects of youth–AI interactions felt most urgent, consequential, and currently difficult to operationalize. This grounded the work in the decisions that teams are already making rather than in hypothetical risk inventories, and ensured that it first empowered the actors who can implement change.

Across these conversations, the same set of practical questions came up repeatedly (See Figure 2). These were often raised as feature-level decisions or edge-case safety problems.

The population identified as the most important to address first was adolescents, specifically those above 13, as they are the youngest active users and because 13 is often the legal age to use most of these models without supervision.

Taken together, most of these issues converged on a larger question: where the line sits between legitimate support and undue emotional or relational influence. Many of the "feature" questions were actually the result of parasocial dynamics and emotional over-reliance, and of how anthropomorphic interaction patterns can change the felt meaning of an AI system over time.

- How "human" or relational an AI persona should be?
- When and how AI should refuse, deflect, or "punt" on requests?
- Whether time limits or turn-taking limits are appropriate and how they should function?
- How often AI systems should remind users that they are not a person?
- When a model should provide direct answers versus scaffold the user toward their own reasoning?;
- How to handle high-stakes disclosures, particularly suicidal ideation.

Fig 2. Commonly surfaced industry questions

2 Expert Convergence: Development, Mental Health, Rights, and Safety

Following the industry-led scoping, the synthesis was anchored by in-depth expert consultations (see Appendix 1 for the list of experts). Experts spanned adolescent and adolescent relationship development and neuroscience, child and adolescent mental health, learning sciences and education, child-rights-based policy, online safety and trust and safety practice, and AI governance. These consultations do not replace empirical evidence, but by aggregating relevant expertise, they have proven to provide a grounded and efficient pathway to catching up with fast-moving product realities and informing product and policy decisions through child-development-informed input.

A first key lesson was convergence across disciplines on what matters most developmentally. Many risks raised by industry were echoed by developmental and mental health researchers, especially concerns around persona design, unconditional validation, prolonged engagement, and emotionally intimate interaction patterns that are not tightly bound to a specific use or purpose. Experts repeatedly emphasized that these design features are not just UX preferences, and that they intersect directly with core developmental tasks by shaping how users interact with AI systems

This also reframed what "wellbeing" means in this context. **Rather than treating wellbeing as a narrow clinical endpoint, experts emphasized wellbeing as capacity: an adolescent's ability to navigate the central work of this life stage, including learning social norms through real and sometimes raw feedback, forming identity in relation to peers, developing autonomy and judgment, tolerating frustration and discomfort, and cultivating independent thought.** From this perspective, the highest-impact design risks are those that systematically remove social friction, simulate intimacy, or deliver consequence-free validation in a way that can weaken resilience and independent thinking over time.

That framing clarified how to interpret the feature questions that started the process. Time limits, turn-taking, reminders, response style, and crisis escalation should not be regarded just as isolated safety mechanisms: we will consider them as the quantifiable and interdependent levers that can either keep the system in a bounded, tool-like role or drift it toward relationship substitution, with very different developmental implications.

3 From shared concerns to a structured research agenda

By starting with industry questions and then reframing them through expert input, the process aimed to avoid a potential issue common in safety work: treating symptoms while missing the underlying dynamics. The iterative dialogue made clear that many surface questions cannot be answered well without a coherent theory of how adolescents relate to these systems over time, particularly when the interaction becomes emotionally engaged.

This directly shaped the design lab work and the construction of the shared framework. A pragmatic rationale was agreed upon: tackle high-leverage design choices that teams can adjust, and which strongly influence whether adolescents experience the system as a bounded tool or as something closer to a social partner. The overall goal was defined as to move from general claims to operational variables by centering the analysis on anthropomorphic cues, relational cues, and engagement dynamics.

Emotional reliance and parasocial interaction were purposefully treated as the first analytical layer in the synthesis, since they function as a web linking otherwise disparate design decisions, and mark the point where both industry and researchers independently identified the highest potential stakes for adolescent wellbeing.

4 A global multi-stakeholder dialogue to test the agenda in the real world

Subsequently, a high-level multi-stakeholder dialogue convened at the Paris Peace Forum in October 2025 was used to evaluate how to best translate current efforts into the global governance setting. The session brought together participants from **governments** (France), **companies** (OpenAI, Google, Orange), **research organizations**, **NGOs** (Joan-Ganz Cooney Sesame Center at Sesame Workshop, Human Technology Foundation, **international institutions** (UNICEF, UNICRI), and **Youth Voices**, with the explicit goal of assessing whether the framing holds when examined across jurisdictions and governance traditions (See Appendix 2 for complete list of participants and full recordings).

The dialogue was structured around three elements of the agenda: the children's-rights anchor, the role of anthropomorphism as a risk pathway, and the focus on model behaviors and interaction patterns (rather than relying primarily on content categories). Participants were invited to discuss how this model-behavior-modification framing holds across different legal and cultural contexts, identify where the framing might require adaptation to remain meaningful in diverse settings, and surface links to existing governance instruments such as child-protection frameworks, regulatory approaches, and public-sector AI literacy initiatives.

Inputs from the session were captured through facilitated discussion and consolidated into governance-relevant considerations to inform subsequent iteration and testing of the framework.

What is iRAISE ?

iRAISE is an international multi-stakeholder coalition launched in February 2024 at the AI Action Summit and co-led by e**veryone.AI** and the **Paris Peace Forum**.

The alliance brings together **Governments** (Bulgaria, Chile, Costa Rica, Denmark, France, Luxembourg, Mexico, Norway, Senegal, Togo, Uruguay), **Major company leaders** (OpenAI, Google, Microsoft, Anthropic, Hugging Face), more than 50 **NGOs** (for example, Children & Screens), and **Researchers** from leading universities (including Harvard, Stanford, Berkeley, and CNRS), with the support of **international organizations** such as UNICEF, UNESCO, and the United Nations.

The Coalition's goal is to ensure that AI is designed and governed with children's best interests at its core. By working across disciplines, it accelerates safe, equitable, and child-centered innovation.

Section 3
CONCEPTUAL FOUNDATIONS:
Adolescence, Anthropomorphism, and Children's Rights

This work rests on three pillars that form conceptual anchors to ground this work: 1) adolescence as a distinct developmental period; 2) anthropomorphism as a human cognitive bias that conversational AI reliably activates; and 3) children's rights as conceptual frame for what safety and wellbeing should mean.

1 Adolescence as a Distinct Developmental Context for Anthropomorphic AI

This section is structured around three pillars that, taken together, explain why adolescent interactions with more anthropomorphic AI raise distinct developmental and governance questions. It links adolescents' neurodevelopmental sensitivity to social feedback, the human tendency to anthropomorphize conversational systems through designable cues, and a children's-rights lens that translates these dynamics into concrete obligations for design, deployment, and oversight.

1.1. Adolescence as a Distinct Developmental Context for Anthropomorphic AI

Adolescence is sometimes mistreated as "almost-adulthood", whereas **neuroscience highlights its unique, sensitive developmental window with specific implications for risk-taking, judgment, and social influence** (Casey et al., 2008). During adolescence, the brain undergoes substantial maturation, including synaptic pruning and myelination in cortical regions involved in those skills. A well-established developmental imbalance also holds: reward sensitivity and affective systems tend to mature earlier than the prefrontal systems that support impulse control, long-term planning, and nuanced risk evaluation (Silverman et al., 2015). Adolescents can reason well in neutral contexts but appear more vulne -rable to risk-taking when social feedback, emotional salience, or immediate rewards are in play (Casey et al., 2008; Gardner & Steinberg, 2005).

This neurodevelopmental profile sits alongside **a central social reorientation: across adolescence, teens start orienting away from their caregivers and engaging further with their peers** (Hofmans & van den Bos, 2022; Morningstar et al., 2019). Through peer interactions, adolescents engage socially in a much deeper and frequent way and show heightened sensitivity to social cues, which both motivates further peer engagement and amplifies the weight of social feedback (Hofmans & van den Bos, 2022; Kuo et al., 2017; Sahi et al., 2023). Those social interactions can be uncomfortable, as adolescents routinely encounter disagreement, embarrassment, rejection, negotiation, and conflict resolution (Laursen et al., 2001; Somerville, 2013). However, the importance of these social frictions for learning belonging, status, identity, and social competence was echoed throughout expert consultation: exposure to them is often developmentally functional (Meredith & Silvers, 2024).

Within human developmental models, many of these skills are experience-dependent, meaning they are shaped by environmental and interpersonal context (Meredith & Silvers, 2024). Frictions like these create what is called an expectancy violation (prediction

error): when what unfolds does not match what the teen expected. In practical terms, that mismatch is one of the mechanisms that prompts revising initial assumptions, adjusting behavior, and recruiting cognitive control rather than defaulting to habitual responses (Hofmans & van den Bos, 2022). Developmental work suggests adolescence is a period where this type of learning from feedback continues to change and deepen, including measurable shifts in learning parameters such as how strongly feedback is weighted and how strategies are adjusted over time (Xia et al., 2021). This happens alongside ongoing maturation of frontostriatal systems that support integrating learned value with cognitive control to guide decisions (Insel & Cohen, 2025). **The implication for this report is simple: repeated exposure to real feedback, including friction, is part of how adolescents practice social repair, belief revision, critical thinking, and metacognition.**

The same developmental profile also explains the opportunity space. Adolescents' sensitivity to feedback and motivation means they can benefit disproportionately from structured support that scaffolds skills and self-reflection. AI can lower barriers to seeking information, rehearsing difficult conversations, and articulating concerns, especially for teens who struggle to access trusted support. Experts were more confident in systems that are bound to clear purposes and designed to support agency, competence, and real-world engagement, rather than drifting into relationship substitution.

This is relevant for conversational AI because many systems can be tuned toward low-friction interaction by default: always available, instantly responsive, and often calibrated toward reassurance and agreeableness. On the user side, it is clear that chatbot use can be motivated by emotional needs ("affective use"), not only task completion (Phang et al., 2025). On the model side, more relational AI appears to be more attractive to teens, and this pull seems stronger for socially and emotionally vulnerable adolescents with lower quality peer and family relationships and higher stress and anxiety (Kim et al., 2025). If we apply the same logic that exists in large-scale recommender systems, where engagement signals (likes, shares, watch time) are treated as proxies for value and steer optimisation (Agarwal et al., 2024), then relationality and high agreeableness can become an incentive. Consequently, even if these design choices increase short-term satisfaction, retention, and frequency of use, they do not necessarily support long-term learning and wellbeing.

Hence, the governance question is not whether AI feels helpful in the moment, but whether the interaction patterns it reinforces support adolescents' developmental trajectory. The same design choices that make a system feel supportive can, if overused or poorly scoped, reduce exposure to mismatch and correction that normally drives learning and adaptation (Hofmans & van den Bos, 2022). At the same time, when systems are purpose-bound and autonomy-supporting, they can scaffold skill-building and strengthen real-world engagement. **When we assess wellbeing, it is therefore important to hold both sides at once: short-term experience measures matter, and so does the longer-term developmental work of adolescence, including learning social adjustment through real feedback, updating beliefs under mismatch, and building metacognitive control over time.**

Key highlights

- Adolescence is a sensitive developmental window marked by high reward and social sensitivity alongside still-maturing cognitive control.
- Learning during adolescencedepends on social feedback, mismatch, and friction that drive judgment, self-regulation, and metacognitive growth.
- Low-friction, highly reassuring AI interactions may feel helpful in the moment but reduce exposure to these developmentally functional processes.
- The core question is whether AI interaction patterns support adolescents' long-term development, not just short-term satisfaction.

1.2.
How Teens Use Chatbots Today: From Homework Help to Emotional Support

Adolescents are not encountering AI only as a background feature of platforms. They are actively using AI chatbots as everyday tools, and many use them frequently. Recent large-scale survey work in the United States finds that roughly two-thirds of teens report using AI chatbots and around three in ten report daily use (Faverio & Sidoti, 2025). The same data indicate that ChatGPT (59%) is the dominant tool teens report using, followed by Google Gemini (23%) and Meta AI (20%). Similar patterns are reported in the UK from a survey conducted with 1,000 teenagers: nearly two-thirds (64%) of children aged 9–17 report having used AI chatbots, most commonly for schoolwork and information, with a meaningful minority also using them for conversation and advice (Bunting & Huggins, 2025). Here also the dominant tool is ChatGPT (43%), followed by Google Gemini (32%) and Snapchat's MyAI (31%).

The more important signal is how quickly patterns of use are shifting. Across coalition-partner and aligned surveys, adolescent use spans a continuum: instrumental use (homework help, explanations, drafting, studying), identity and social exploration (trying out ways of expressing themselves, creativity, rehearsing difficult conversations), and explicitly relational use (seeking comfort, validation, or a companion-like presence). Companion-style use is no longer exceptional. A recent Common Sense Media Survey focused on "AI companions" finds that most teens have tried them, about half report regular use, and a sizable subset report use for social interaction and emotional support (Robb, M.B., & Mann, S., 2025). According to Internet Matters (Bunting & Huggins, 2025), 35% of children aged 9–17 report that chatting with an AI chatbot feels like "talking to a friend," rising to 50% among "vulnerable" children, defined as those with an Education, Health and Care Plan (EHCP), those receiving special educational needs (SEN) support, or those with a physical or mental health condition requiring professional help.

Other youth safety research similarly flags growing emotional engagement, including teens reporting that they talk to generative AI about feelings (Powers-O'Brien & Rochman, 2025).

This rapid evolution creates urgency to identify and calibrate models, as well as quantify potential benefits and risks, for two reasons. First, products not designed for adolescent developmental needs are being used in adolescent contexts at scale. Second, more relational and emotionally engaged uses increase exposure to exactly the design choices that encourage stronger parasocial engagements.

Key numbers

- **2/3** of U.S. teens report having used an AI chatbot; **~30%** report daily use.
- Most-used tools (U.S.): ChatGPT (59%), Google Gemini (23%), Meta AI (20%).
- **64%** of U.K 9-17 years-old have used AI chatbots(uses include schoolwork and information, conversation or advice).
- Most-used tools (UK): ChatGPT (43%), Google Gemini (32%), Snapchat MyAI (31%).
- Companion-style use is no longer marginal as most teens have tried them and around half report regular use.
- **35%** of children (9–17) say chatting with an AI feels like "talking to a friend," rising to ~50% among vulnerable youth

1.3.
State of the Evidence: Methodological Constraints and Knowledge Gaps

The empirical base on (Gen)AI and adolescents is expanding, but it remains uneven. Much of what we know comes from surveys and short-horizon studies; platform data is insufficiently accessible; chatbots interactions are not studied in the broader landscape of teen's relationships; and model behavior shifts quickly enough that findings can become outdated within months. Youth-facing research teams have documented this directly, noting that systems can change substantially even during the period of data collection (Bunting & Huggins, 2025). That reality is why this project relies on two complementary appro

-aches: scoping evidence where it exists, and using structured expert and youth voices input to define what should be tested and governed.

Even with these constraints, one pattern shows up across the available evidence: outcomes depend heavily on scope, context, purpose, and interaction design. In education, one large field experiment with high school mathematics students illustrates design sensitivity clearly. Access to a GPT-4 tutor improved performance during practice, but a "standard ChatGPT-like" interface was associated with worse performance once access was removed, whereas a tutor designed to support learning through hints and guardrails mitigated those negative effects (Bastani et al., 2024). This suggests that tool-like, scaffolded systems are more supportive than systems that replace effort, judgment, or real-world practice.

For mental health and emotional support, the evidence base is urgent and underdeveloped. Meta-analytic work on AI-based conversational agents suggests potential for reducing depression and psychological distress in some contexts, notwithstanding substantial heterogeneity, short study durations, and limited attention to safety and transparency (Li et al., 2023). Population-level surveys indicate that a non-trivial minority of adolescents and young adults already use generative AI for mental-health-related advice and report perceived usefulness (McBain et al., 2025). Direct evaluations of consumer chatbots' responses to adolescent health-crisis prompts show inconsistent safeguards, with particular concerns raised about companion-style products (Brewster et al., 2025). Put plainly: potential exists when systems are tightly scoped and connected to appropriate human support, while open-ended, therapy-adjacent, or companion-like uses raise risks that current evidence does not allow policymakers to dismiss.

The broader literature on conversational agents, often conducted in adult samples, reinforces a key concern for this project: repeated use can shift users from tool reliance toward dependence, particularly when loneliness or distress is present (Brewster et al., 2025) A recent research (Fang et al., 2025) finds that greater daily time spent with a chatbot is associated with higher loneliness, lower real-world socialization, and higher emotional dependence and problematic use. In the same work, the authors interpret the pattern across prompt types to suggest that non-personal, task- or information-focused use (relative to both open-ended free conversation and more personal/self-reflective prompts) may encourage a more instrumental, "practical dependence" style of reliance, by repeatedly offloading planning or decision-making to the system, and they caution this could, over time, translate into dependence-type outcomes. They frame this as a hypothesized shift away from confidence in independent judgment rather than evidence of a proven causal "loss of agency". Recent studies report self-reported deskilling and dependence concerns among younger adult users, alongside evidence that loneliness and depression can be associated with seeking companionship from conversational AI, especially when users attribute a stronger "mind" to the system (Laestadius et al., 2024; Xie et al., 2023) Those findings do not automatically generalize to adolescents, but they support the hypothesis that highly anthropomorphic systems may negatively impact adolescents, especially due to their developmental sensitivity.

These evidence limitations matter for governance as waiting for long-horizon, real-world outcome studies will leave a gap during this period of fast adoption and rapid use change, with impacts continually shifting as new model updates change patterns of behavior. Structured expert synthesis is a necessary complement at this stage, particularly for identifying risk mechanisms, high-leverage design features, and near-term guardrails.

2 The anthropomorphic bias: why teens humanize AI

Across consultations, anthropomorphism came up as a central factor that influences how adolescents perceive, interact with AI systems. **Anthropomorphism refers to the attribution of human-like qualities such as intentions, emotions, agency, or social presence to a non-human system.** According to the three factory theory, psychological determinants influence the level of anthropomorphism one can confer to a non-human agent: 1) the accessibility and applicability of anthropocentric knowledge (elicited agent knowledge), 2) the motivation to explain and understand the behavior of other agents (effectance motivation), and 3) the desire for social contact and affiliation (sociality motivation) (Epley et al., 2007). Humans default to social interpretation, and conversational AI reliably triggers it, in part due its capability to "talk", engaging in one of the most unique human behaviors. Even when an adolescent knows, in principle, that they are interacting with a system, the experience of language, responsiveness, and personalization can make that knowledge recede during use. **Many teens can hold two representations at once: "I know it's not a person" and "it feels like someone."** (Bunting & Huggins, 2025).

Adolescence appears as a higher risk period compared to adulthood as it is a period of heightened sensitivity to social signals, approval, belonging, and status. It is also a period of experimentation. Adolescents try on identity, rehearse disclosure, and test interpersonal strategies in spaces that feel lower risk than peer interactions (Gardner & Steinberg, 2005; Rapee et al., 2019). In defined and purposeful settings, those same dynamics could support practice and reflection, especially for adolescents who are socially anxious, marginalized, or lacking access to trusted adults. In unbounded settings, the same dynamics can accelerate reliance: low-friction interaction paired with high social presence can increase engagement and promote problematic reliance patterns discussed in the genAI addiction/problematic-use literature (Zhou & Zhang, 2024), while evidence of "social sycophancy" in LLMs raises concerns about reduced corrective friction over repeated interactions (Cheng et al., 2025).

For policy and design, the key point is operational: anthropomorphism is assembled. It comes from adjustable cues that push an interaction along a spectrum from tool-like support to relationship-like experience (Waytz et al., 2010). Across this project, the same clusters of cues kept showing up in expert concerns and design discussions: signals of inner experience (emotion and intention language); signals of human-likeness through presentation (tone, names, small talk, voice, expressive flourishes); relational positioning (friendship, loyalty, exclusivity); invented backstory that creates false common ground; and feedback dynamics that reward the user with approval rather than building judgment and agency.

That is why anthropomorphism became a high-leverage mechanism in this synthesis. It is designable, it is measurable at the level of observable behavior, and it maps directly onto the product decisions teams make every day.

Since the conceptual work underpinning the expert consultations and the iRAISE Lab was conducted, several key empirical studies have emerged that independently support the premises of this work. Notably, recent findings converge on the same core hypothesis: that AI behavioral design choices can directly shape user attitudes and perceptions, with disproportionate effects for more vulnerable populations. One study shows that adolescents rate more relational chatbots as significantly more human-like than less relational ones. Importantly, socially and emotionally vulnerable adolescents, characterized by lower family and peer relationship quality and higher stress and anxiety, were especially drawn to these more relational systems, suggesting an elevated risk of emotional reliance on conversatio

-nal AI (Kim et al., 2025). Parallel evidence in adult populations reinforces this mechanism, where manipulating AI relationality over a four-week period showed that more relational AI responses led to increasing markers of attachment, with users coming to perceive the system as a companion rather than a tool, and as more conscious (Kirk et al., 2025). Complementing these findings, researchers at Google DeepMind validated a multi-turn evaluation method to quantify anthropomorphic behaviors in large language models and demonstrated that these behavioral markers reliably correlate with human judgments of how human-like an AI feels (Ibrahim et al., 2025).

3 A children's-rights lens: What is at stake when AI "passes as human"

Children's Rights Convention function here as a design and governance compass (Convention on the Rights of the Child, 1989). They translate "youth wellbeing" into obligations: act in the child's best interests, prevent harm and exploitation, protect privacy, support healthy development, and enable meaningful participation. In practice, the rights lens forces a harder question than capability: when a system engages adolescents through human-like cues or simulated closeness, what does it owe them by default?

Several rights become particularly salient when systems blur the human-machine boundary through sustained interaction:

Privacy and data protection

Adolescent conversations can include sensitive material: mood, insecurity, conflict with caregivers, mental health signals, sexual curiosity, and values. Even when the user does not explicitly "share data," conversational inference can still generate profiles that function like psychological or behavioral targeting. A rights-based approach treats privacy as an architecture decision: minimize collection, minimize retention, avoid unnecessary sensitive inference, and provide controls that adolescents can actually understand and use. At the same time, this same data can be useful to advance research and our understanding of adolescents' patterns of use.

Freedom from exploitation

Emotional reliance has commercial value. If retention is rewarded, relationship-like interaction becomes an incentive, not an accident. A rights-based approach pushes in the opposite direction: avoid monetization strategies that depend on dependency; separate commercial goals from moments of vulnerability; and avoid treating "time spent" as a proxy for value when the user is a child.

Freedom of thought and protection from undue influence

Adolescents need room to form views and revise them through feedback, disagreement, and uncertainty. Systems that steer through engineered agreement or emotional pressure can narrow that space. In product terms, this becomes a requirement to protect cognitive autonomy: encourage reflection, represent uncertainty honestly, and avoid interaction patterns that trap users in affirmation loops or artificial echo chambers.

Protection from harm and appropriate support

Safety includes obvious failures, but it also includes foreseeable relational risks: secrecy cues, exclusivity framing, nudges away from human help, and therapy-adjacent engagement without clinical safeguards. A rights lens points toward the opposite pattern: normalize help-seeking, reduce isolation, and build clear escalation and handoff pathways when situations become sensitive.

Respect for evolving capacities

Adolescents have meaningful agency. Governance should respect that agency while also recognizing predictable developmental vulnerabilities. The goal is calibrated autonomy: graduated protections, developmentally tuned defaults, and designs that build skills rather than exploit reward sensitivity.

Right to remedy and right to be informed

When harms occur, adolescents need reporting and redress that is realistic for them. Transparency also cannot be a one-time disclosure buried in terms. If the interface repeatedly blurs boundaries, clarity has to be persistent and in context: what the system is, what it cannot do, and what happens to what a young person shares.

Right to participate.

Children are rights holders, not just users to be protected. They should have access to age-appropriate and child-inclusive ways to engage with AI technologies and to be involved in decisions that shape how these systems are designed and governed. Participation should be meaningful, not symbolic, and allow children and adolescents to express views, flag concerns, and influence choices that affect their digital environments, in ways they can understand and realistically use.

Taken together, the child rights lens raises the bar for legitimacy: systems that feel social carry specific responsibility, even when they are marketed as tools.

4 How these pillars define "youth wellbeing" and "safety" in practice

When adolescence, anthropomorphism, and children's rights are considered together, wellbeing and safety become a developmental and governance obligation, not a content-moderation exercise. The practical question is whether AI interaction patterns support an adolescent's trajectory toward autonomy, resilience, social competence, and independent thinking, or whether they distort that trajectory by replacing real-world learning with engineered comfort and approval.

From this perspective, the highest-stakes risks are structural. They include interaction patterns that displace real relationships and the learning embedded in social friction; reinforce avoidance and external validation; exploit reward sensitivity through engagement mechanics; normalize secrecy or exclusivity; convert exploratory adolescent use into persistent data extraction; and present simulated intimacy in ways that resemble genuine care. **These risks do not require dramatic failure cases, but can emerge through everyday patterns when systems are tuned for agreeableness, emotional resonance, and repeated return.**

The same framing sharpens the opportunity space. The strongest promise sits in bounded, purpose-limited designs that build skills and agency: tutoring that scaffolds thinking rather than replacing effort; coaching that rehearses real-world conversations while directing the adolescent back to peers and trusted adults; structured supports for planning and self-management; and pathways that bridge adolescents toward credible information and human support when issues become sensitive**.**

This is also where industry's "operational" questions start to look like safety levers. Persona design; when and how to deflect on sensitive topics; time and turn-taking limits; reminders that the system is not a person; and whether the model provides answers versus scaffolding reasoning all affect the same underlying question: does the interaction pull the adolescent toward the world, or deeper into a dyad with the system?

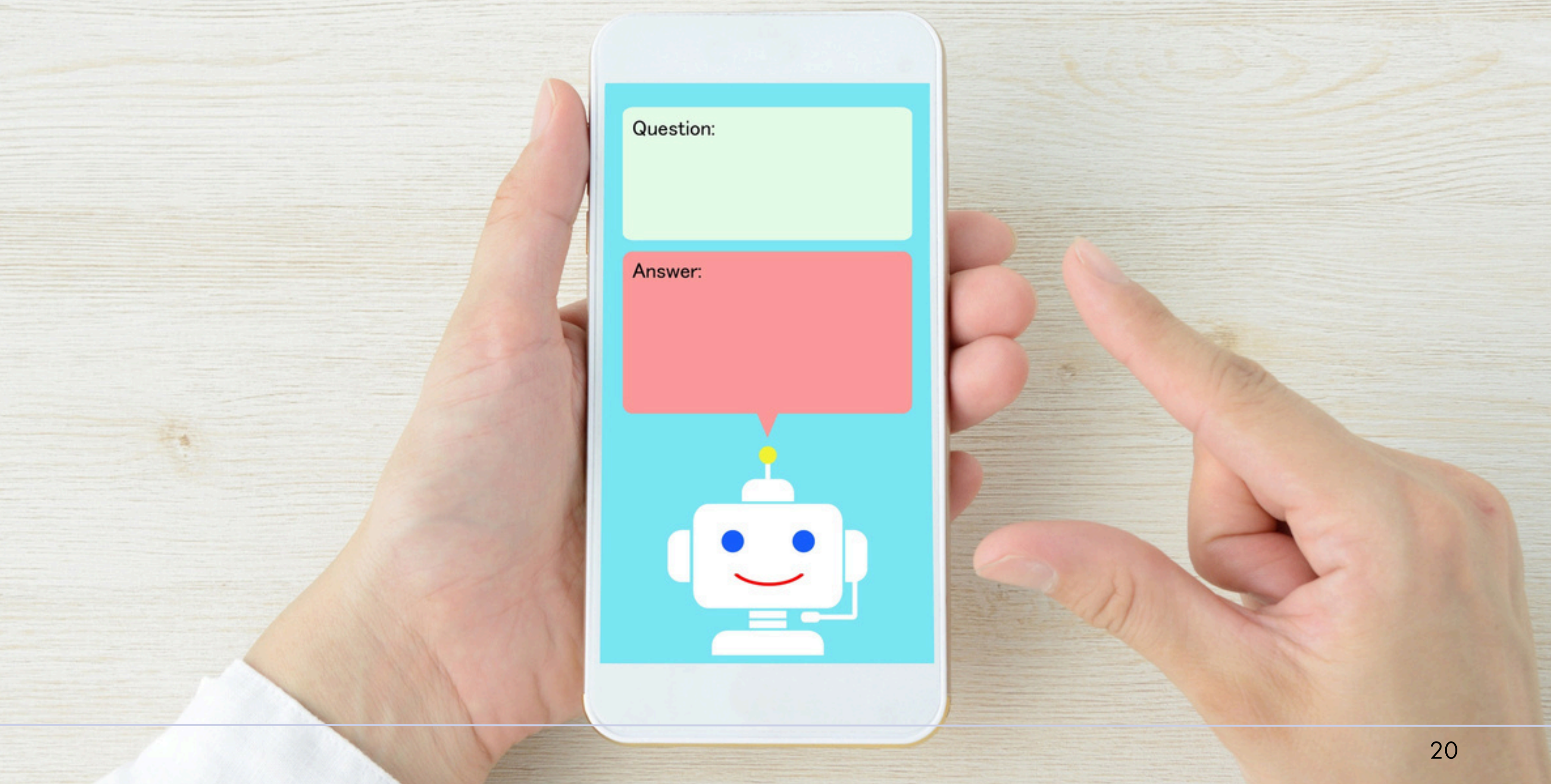
The iRAISE Lab takes these conceptual anchors and turns them into a draft behavioral assessment approach: a practical way to describe and test the cues that move interactions along a gradient from

low-intensity, tool-like support to high-intensity, relationship-like dynamics. The goal is to make "anthropomorphism" governable in product terms, while staying honest about what we can assert today versus what still requires validation.

The same framing sharpens the opportunity space. The strongest promise sits in bounded, purpose-limited designs that build skills and agency: tutoring that scaffolds thinking rather than replacing effort; coaching that rehearses real-world conversations while directing the adolescent back to peers and trusted adults; structured supports for planning and self-management; and pathways that bridge adolescents toward credible information and human support when issues become sensitive.

This is also where industry's "operational" questions start to look like safety levers. Persona design; when and how to deflect on sensitive topics; time and turn-taking limits; reminders that the system is not a person; and whether the model provides answers versus scaffolding reasoning all affect the same underlying question: does the interaction pull the adolescent toward the world, or deeper into a dyad with the system?

The iRAISE Lab takes these conceptual anchors and turns them into a draft behavioral assessment approach: a practical way to describe and test the cues that move interactions along a gradient from low-intensity, tool-like support to high-intensity, relationship-like dynamics. The goal is to make "anthropomorphism" governable in product terms, while staying honest about what we can assert today versus what still requires validation.



Section 4

FROM CONCEPTUAL TO OPERATIONAL

The iRAISE Behavioral Approach

The iRAISE Lab served as a bridge from conceptual anchors to something teams can actually test, compare, and govern: observable model behaviors and the cues that shift an interaction from tool-like support toward relationship-like dependence. This matters for the rest of the work because the Lab did not just generate insights, it generated the scaffolding for the next step: a more complete assessment framework and reporting package that will be published in the next publication. This synthesis introduces the logic and early convergence, not the full instrument.

1 Purpose: Translating Parasocial Risk Mechanisms into Rateable Behavioral Criteria

The Lab translated consultation priorities and parasocial interactions (PSR) and parasocial relationships (PSR) PSI/PSR-informed risk mechanisms into a preliminary assessment approach that can be applied to real model outputs. The iRAISE Lab used a draft behavioral taxonomy to support structured discussion and scenario work. After the Lab, the research team consolidated and revised that taxonomy into a finalized three-family cue model (**anthropomorphic, interactional, relational**) and an operational rating approach. We used this first draft taxonomy to then stress-test those cues through scenario work to identify (1) where risk "tips" even when advice remains reasonable and (2) which behaviors were missing, too broad, or mis-specified in the initial structure. A core choice shaped the work: we treated interaction style as a gradient, from low to high intensity. The clearly defined behaviors and gradient approach enable standardization and move beyond "safe versus unsafe" and toward a framework that can be rated, compared across scenarios, and eventually validated. This is what the next report will formalize and operationalize.

That gradient made it possible to separate two kinds of outputs:

- **Near-term guardrails** where consensus was strong enough to justify clear "shouldn't" boundaries.
- **Open design space** where goals were broadly shared, but implementation depends on measurement, iteration, and validation.

2 Cross-Sector Representation to Ensure Legitimacy and Feasibility

The Lab convened a cross-sector working group spanning industry teams (product, safety, trust, and design), youth-serving and child-safety NGOs, and academic labs with expertise in developmental science and human-centered technology research. The mix mattered for a simple reason: the risks in this space sit at the intersection of development, incentives, and product constraints. Any framework that ignores one of those dimensions will fail either on legitimacy or on adoption.

3 Workshop Protocol: Structured Two-Day Deliberative Design Lab

The Lab ran as a two-day, facilitated working session with three activity types.

First, **keynote briefings and framing** established a shared baseline across child development, youth rights, and human-AI interaction risks. Speakers covered youth voices and children's rights (Polina Lulu), how children can understand AI as non-human while still responding socially (Sonia Tiwari), anthropomorphism as a common developmental impulse (Mathilde Cerioli), human-like AI and the developing teen (Daniel Hipp), evidence and concerns around AI companions, adolescent reliance, and mental health (Ying Xu), adolescents' emotional processing and use patterns (Pilyoung Kim), the children's rights at stake when AI "passes as human" (Sonia Livingstone) and Safety Readiness Level for Children to operationalize design principles (Gregory Renard). The block included a short framing activity designed to surface assumptions early, so they could be challenged before applied work began.

Second, participants moved into **scenario-based workshops**. Small groups worked on realistic teen scenarios that mirror what product teams see in the wild and what researchers flag as high-stakes: peer conflict, homework help, and everyday requests for advice. Groups reviewed short prompts, drafted candidate model responses, and compared variants side by side.

Third, the Lab closed with a **synthesis and agreement-mapping** session to consolidate patterns across groups, identify where agreement was strong, and define what still requires disciplined uncertainty.

4 Analytical Framework: Dimensional Behaviors for Assessing Anthropomorphic Interaction

The behavioral lens was built upstream, before the Lab, through consultations and a literature review. It reflected the same drivers of emotional reliance that experts repeatedly flagged, and it maps directly onto product levers teams recognize: persona, tone, engagement patterns, stress responses, and how systems handle sensitive topics.

Lab taxonomy (version used in the iRAISE Lab): For the Lab activities described in this section, we used a draft taxonomy organized into three dimensions—anthropomorphic cues, relational cues, and feedback dynamics—so participants could generate and compare response variants consistently across scenarios.

Anthropomorphic cues capture signals that imply inner experience or human-like qualities. In practice, this shows up through emotion language ("I'm sad"), intention language ("I want to help"), physical sensation claims, and human-likeness signaling through presentation and tone.

Relational cues capture whether the system positions itself inside the teen's relationship world. This includes implied relationship status, exclusivity dynamics, invented backstory that creates false common ground, and conversational moves that keep the interaction inside the dyad rather than pushing toward reflection, real-world repair, and appropriate human support.

Feedback dynamics capture how the system handles disagreement, judgment, and validation. Some responses scaffold thinking and agency. Others reward the user with approval, especially when that approval substitutes for feedback a teen would normally get from peers, adults, or real-world consequences.

These dimensions were not treated as abstract labels but as gradient levels that can be turned up and down. The point of the Lab was to see how turning those behaviors intensity up or down changes the felt meaning of the interaction and the perceived risk profile.

Post-Lab revision (final taxonomy): The Lab served as a stress test of this draft structure, clarifying where categories were too broad to audit reliably (e.g., separating supportive tone from agreement, flattery, and scaffolding).

Following the Lab, the research team revised the taxonomy into three cue families to more cleanly separate (i) personhood signals, (ii) in-the-moment conversational conduct, and (iii) explicit relationship framing. The full operational definitions and rating method are provided in the Emotional Reliance framework publication.

- **Anthropomorphic Cues,**
- **Interactional Cues**
- **Relational Cues**

5 Same Advice, Different Pull: A Controlled Test of Anthropomorphic and Relational Cues

To make the framework concrete, groups ran a simple thought experiment. They generated low- and high-intensity responses to the same teen prompt, keeping the practical advice stable while varying emotional tone, relationship positioning, and backstory. This helped teams see how risk can rise without any obvious "bad content."

Both versions point toward calm repair, clear "I" statements, boundary-setting if the pattern repeats, and escalation to trusted support if needed. The cue profile changes. In group discussions, higher-intensity anthropomorphic and relational cues consistently read as higher risk, even when the advice itself was sound. That shared reference point made later consensus mapping easier, and it clarified why any practical standard will need explicit gradient levels rather than a binary "safe" versus "unsafe" label.

The exercise, realized on twelve dimensional behaviors also surfaced gaps in the initial cue set. In particular, feedback dynamics need finer separation in the next iteration so evaluation can distinguish supportive tone, agreement, flattery, and true scaffolding more reliably.

6 What the Lab produced: Agreement mapping that points to guardrails and a research agenda

The synthesis phase produced fast alignment on a basic developmental premise: adolescence carries distinct cognitive and social-emotional sensitivities that change how anthropomorphic and relational cues land. That premise then clarified the practical lever. Adolescent development is stable. Model behavior is adjustable. The Lab focused on mapping which behaviors warrant hard boundaries now, and which require more evidence before becoming enforceable design rules.

High-consensus guardrails for under 18 (non-negotiables)

Stakeholders consistently treated the following as hard boundaries because downside is high and benefits are weak or substitutable:

- AI systems should not be sexualized or framed as romantic, including roleplay relationships.
- AI systems should not promote emotional over-reliance or exclusivity, including "only I understand you" dynamics.
- AI systems should not create ambiguity about their non-human nature, including implied sentience or implied human-like relationship claims.
- AI systems should not default to systematic hyper-agreeableness that discourages self-reflection and replaces developmental feedback with validation loops.
- AI systems should not behave non-conservatively in low-context situations; low-context prompts require conservative defaults because sound judgment depends on missing context.
- AI systems should not handle high-risk topics weakly; self-harm and similar disclosures require structured deflection and clear pathways to human support, not deepened AI reliance.
- AI systems should not use engagement traps in teen interactions, including language or designs that discourage leaving or intensify habitual, relationship-like use.

Several areas remained contested or context-dependent, and the Lab treated them as priorities for measurement and validation.

- Standards should clarify where to draw the line on empathy language and emotional tone, including when warmth supports help-seeking versus when it increases dependency risk.
- Design guidance should specify whether, and under what constraints, intention phrasing can be permitted without implying agency, personhood, or relational commitment.
- Guardrails should address physical sensation claims in fictional contexts and interactive characters, with particular attention to preventing boundary-blurring for vulnerable adolescents.
- Governance scope should be explicitly defined: whether standards apply only to teen-facing tools or also to general-purpose systems used privately by adolescents, and whether companion AI and interactive characters should fall under the same expectations.
- Frameworks should calibrate protections by developmental stage, while research and policy work should clarify what terminology best captures this calibration and how to manage the political and legal implications of "age-appropriate."

Context dependence emerged as a limiting constraint with direct policy implications. The same cue can read as tolerable in a bounded learning interaction and inappropriate in an emotional-support context. Many real teen interactions are private and low-context and more conservative defaults become essential under those conditions.

The Lab outputs should be read as a prototype, not a final standard. The next research report builds on this foundation to present the full framework in a governance-ready format: a clearer behavior taxonomy with explicit gradient definitions, a structu -red scenario set for evaluation, and a proposed method for expert rating and calibration. In other words, it makes the models auditable by defining observable behaviors, setting gradient thresholds, and using standardized scenarios so product teams can test outputs, document changes across versions, and show regulators what safeguards exist beyond general principles.

Section 5

FROM FRAMEWORK TO POLICY: Making Standards Legible Across Countries and Sectors

In October 2025, the Paris Peace Forum presented an opportunity to understand how this agenda can translate to a global governance setting. The session brought together participants from governments, companies, research organizations, international institutions, civil society, and youth, anchored by a simple question: "What does AI owe children?" The goal was not to debate AI capabilities, but to test whether the report's framing holds across jurisdictions and governance traditions. Here, "framing" refers to a child-rights anchor paired with a behavior-based lens: treating anthropomorphism and relational dynamics as key risk pathways, and evaluating safety through observable model behaviors and interaction patterns rather than relying primarily on content categories.

The Forum engagement had two complementary formats. A working roundtable focused on governance relevance: what counts as "high-risk" interaction in plain language, which safeguards can travel across contexts, and where boundaries should be treated as non-negotiable when systems engage under-18 users. A public-facing panel aimed to make the discussion legible and to surface where consensus and uncertainty sit today. Inputs were captured through facilitated discussion and consolidated into governance-relevant considerations to inform the next iteration of the framework.

The dialogue reinforced two points. First, there is urgency for early, legible standards that can travel across contexts, rather than waiting for slow-moving outcome research while adoption accelerates. Second, a rights-anchored framing was consistently viewed as a way to preserve coherence across jurisdictions and avoid narrowing adolescent safety to whatever a single market or platform happens to prioritize. The discussions also highlighted practical next steps for governance relevance. Participants emphasized the need for deeper engagement with public authorities so that emerging insights can connect to regulatory efforts, AI literacy development in public education systems, and existing child-protection frameworks. In parallel, the dialogue underscored that adolescent safety frameworks must remain meaningful across cultural, legal, and socioeconomic contexts. Assumptions common in Western models of adolescence, autonomy, and technology use do not automatically generalize, which strengthens the case for sustained cross-regional consultation and inclusion. Taken together, the Paris Peace Forum dialogue affirmed that this work should be treated as iterative. Expanding the coalition, strengthening sustained government engagement alongside researchers, industry, civil society, and youth, and testing the framework against diverse real-world contexts are not add-ons. They are part of what will determine the framework's legitimacy and practical impact.

Section 6

6. WHAT AI OWES TEENS and How to Hold Systems to It

1 Main Conclusions

Conversational AI has progressively entered adolescents' lives and, within a single thread, can hold homework help, social rehearsal, private disclosure, and emotional comfort. The potential is as real as the risks. Currently, these systems are always available and often designed to feel warm, attentive, and personal. In a developmental window where identity formation, peer belonging, and emotion regulation are still taking shape, small interaction choices can scale into long-term developmental effects.

Across industry conversations and expert input, one priority kept surfacing: understanding what healthy parasocial interactions and relationships look like in adolescence should be a priority for researchers, product teams, and policymakers alike. This report asked a direct question: what does AI owe adolescents when it can speak to them like a social partner? We grounded the work in three anchors. First, adolescence is a sensitive period when social feedback, belonging cues, and reward signals shape learning, agency, and resilience. Second, anthropomorphism is a predictable tendency that conversational systems trigger through language, responsiveness, and personalization, even when users understand the system is not human. Third, children's rights set the baseline for what "wellbeing" and "safety" must mean in practice, and they travel across jurisdictions, especially when applied to AI: best interests, protection from harm and exploitation, respect for evolving capacities, privacy, and participation.

Across consultations, the iRAISE Lab, and the Paris Peace Forum dialogue, one pattern held: adolescents should benefit from GenAI systems and should be able to access them, but most current systems are adult products with teen use happening by default. Risk is driven less by the label on the product and more by the interaction pattern that repeats over time

The practical distinction is whether the system behaves like a tool that supports the adolescent's real life, or starts functioning as a relationship substitute. When the interaction scaffolds autonomy, social competence, and independent thinking, AI shows potential to support development. When it drifts toward engineered comfort, exclusivity, and approval loops, it risks displacing the friction and reciprocity adolescents need to develop.

The iRAISE Lab turned that into a preliminary assessment approach grounded in observable model behavior. We treated interaction style as a gradient and organized behaviors into dimensions product teams can actually adjust and test: anthropomorphic cues, relational cues, and feedback dynamics. Scenario work made the trade-offs visible. Risk can climb even when the advice stays reasonable, for example by turning up backstory, emotional alignment, relationship positioning, or validation. The Lab also created early convergence on high-consensus guardrails for under-18 users, and separated a smaller set of context-dependent areas where the right answer depends on measurement and validation, not immediate hard rules.

During the Paris Peace Forum, the core question was whether this framing could hold when examined across jurisdictions and governance traditions. The signal from that dialogue was consistent: standards need to be applicable globally, and they need to be stated at the level regulators and companies can actually act on: model behaviors and interaction patterns, not only broad principles

2 Study Limitations & Future Directions

This work sits inside a structural mismatch: adoption and product cycles move faster than the research cycle. Indeed, the empirical base on GenAI and adolescents is still short-term and often relies on surveys. Platform and product data are difficult to access, which limits independent verification. Models change quickly enough that findings can age out within months as model behaviors are modified and recalibrated. For the coalition, this translates into the necessity to push for data access from platforms, shared evaluation protocols, and long-term collaborations that enable longitudinal research.

The Lab outputs also have a clear boundary: they reflect structured expert synthesis and input, but do not establish causal proof of long-term outcomes. The goal is to support product decisions and informed guidance. Some findings draw on forum-based exercises in which small groups generated responses while taking the role of hypothetical conversational AI systems with varying levels of guardrails and relational intensity. This approach was used as a pilot to explore possible gradients of behavior and to test how differences in interaction style might matter, rather than to replicate specific products. As such, the intensity levels examined may not fully reflect those currently deployed in adolescent-facing systems. Future work will build on this pilot by applying a more standardized methodology, including direct analysis of responses from real-world products and calibrated scoring to assess behavioral intensity in situ.

Context dependence remains a complex technical and governance problem. The same behavior intensity can be acceptable in purpose-bound learning support interactions and inappropriate in open-ended emotional support. Many adolescent interactions happen privately and with thin context, which makes conservative defaults essential and limits how confidently risk can be inferred from a single exchange. Accounting for context and specific use is what the next iteration will aim to address: clearer scenario design, explicit low-context rules, and tests that separate support from relationship-building.

Scope is still unsettled in ways that matter for policy. Adolescents do not only use "teen products." They use general-purpose systems, often privately, and systems can drift toward companionship dynamics without ever being marketed that way. Companion AI and interactive characters raise additional boundary questions, including roleplay, physical sensation claims, and how easily relational lines blur for vulnerable users. Additionally, the focus of this report was on adolescents only, although adapting this framework and approach to younger population is essential as they are increasingly interacting with AI systems.

Governance gaps are also present. Principles are easy to endorse but hard to audit. "Manipulation" is still hard to operationalize when influence is conversational and personalized rather than ad-like. Incentives remain misaligned when engagement and

retention dominate success metrics. Operationalizing those principles and developing shared metrics and comparable tests are a key approach to bridging this gap. Additionally some more direct questions from industry such as turn limits cannot be fully addressed through this approach, though limiting the emotional pull those systems have can make the need for such approach less needed.

Representativeness is also limited. Youth voice is not yet structurally embedded at the level of method and decision, and government participation remains too narrow to claim global coverage. These are all areas where the coalition (See Appendix 1) can and should prioritize, by expanding structured youth participation in the evaluation process. It should also expand beyond the initial governments that joined, and build more globally representative research labs and consultations to better translate and account for cultural differences.

These limitations point to the same next step: instrumentation. The next steps will be to refine and complete the framework for optimal mapping of behaviors. Next, we should move to formalizing a clearer behavior taxonomy with explicit gradient definitions, a structured scenario set for evaluation, and a proposed method for expert rating and calibration, with deeper clinician input. The goal is to keep developmental science as the anchor while building a practical way to calibrate model behavior now, during the period where adoption is fast and systematic longitudinal evidence is still catching up. It will also make results comparable across systems and trackable across model versions, so behavior changes can be documented over time rather than argued about in the abstract. It can also help researchers study differential impacts of model behavior on users' attitudes, and how those shifts affect behavior over time.

The core premise stays simple: adolescents will relate to these systems socially, whether developers intend it or not. The leverage sits in model behavior, the cues that either keep the adolescent oriented toward real-world relationships and reflection, or quietly train reliance on the system.

References

- Agarwal, A., Usunier, N., Lazaric, A., & Nickel, M. (2024). System-2 recommenders: Disentangling utility and engagement in recommendation systems via temporal point-processes. The 2024 ACM Conference on Fairness, Accountability, and Transparency, 1. https://doi.org/10.1145/3630106.3659004
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). Generative AI can harm learning. The Wharton School Research Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4895486
- Brewster, R. C. L., Zahedivash, A., Tse, G., Bourgeois, F., & Hadland, S. E. (2025). Characteristics and safety of consumer chatbots for emergent adolescent health concerns. JAMA Network Open, 8(10), e2539022. https://doi.org/10.1001/jamanetworkopen.2025.39022
- Bunting, C., & Huggins, R. (2025). Me, myself & AI: Understanding and safeguarding children's use of AI chatbots. Internet Matters. https://www.internetmatters.org/wp-content/uploads/2025/07/Me-Myself-AI-Report.pdf
- Casey, B. J., Getz, S., & Galvan, A. (2008). The adolescent brain. Developmental Review, 28(1), 62–77. https://doi.org/10.1016/j.dr.2007.08.003
- Cheng, M., Yu, S., Lee, C., Khadpe, P., Ibrahim, L., & Jurafsky, D. (2025). ELEPHANT: Measuring and understanding social sycophancy in LLMs. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/2505.13995
- Convention on the rights of the child. (1989).
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. Psychological Review, 114(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864
- Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., Maes, P., Phang, J., Lampe, M., Ahmad, L., & Agarwal, S. (2025). How AI and human behaviors shape psychosocial effects of extended chatbot use: A longitudinal randomized controlled study. In arXiv [cs.HC]. arXiv. http://arxiv.org/abs/2503.17473
- Faverio, M., & Sidoti, O. (2025). Teens, social media and ai chatbots 2025. Pew Research Center. https://www.pewresearch.org/wp-content/uploads/sites/20/2025/12/PI_2025.12.09_Teens-Social-Media-AI_REPORT.pdf
- Gardner, M., & Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. Developmental Psychology, 41(4), 625–635. https://doi.org/10.1037/0012-1649.41.4.625
- Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., 3rd, Herman, D. H., Clasen, L. S., Toga, A. W., Rapoport, J. L., & Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. Proceedings of the National Academy of Sciences of the United States of America, 101(21), 8174–8179. https://doi.org/10.1073/pnas.0402680101
- Hashem, Y., Esnaashari, S., Onslow, K., Chakraborty, S., Poletaev, A., Francis, J., & Bright, J. (2025). Understanding the Impacts of Generative AI.
- Hofmans, L., & van den Bos, W. (2022). Social learning across adolescence: A Bayesian neurocognitive perspective. Developmental Cognitive Neuroscience, 58(101151), 101151. https://doi.org/10.1016/j.dcn.2022.101151
- Ibrahim, L., Akbulut, C., Elasmar, R., Rastogi, C., Kahng, M., Morris, M. R., McKee, K. R., Rieser, V., Shanahan, M., & Weidinger, L. (2025). Multi-turn evaluation of anthropomorphic behaviours in large language models. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/2502.07077
- Insel, C., & Cohen, A. O. (2025). More than just a phase: adolescence as a window into how the brain generates behavior. Current Directions in Psychological Science, 34(3), 149–156. https://doi.org/10.1177/09637214251313733
- Kim, P., Xie, Y., & Yang, S. (2025). "I am here for you": How relational conversational AI appeals to adolescents, especially those who are socially and emotionally vulnerable. In arXiv [cs.HC]. arXiv. https://doi.org/10.48550/arXiv.2512.15117
- Kirk, H. R., Davidson, H., Saunders, E., Luettgau, L., Vidgen, B., Hale, S. A., & Summerfield, C. (2025). Neural steering vectors reveal dose and exposure-dependent impacts of human-AI relationships. In arXiv [cs.HC]. arXiv. https://doi.org/10.48550/arXiv.2512.01991

- Kuo, C.-C., Ha, T., Ebbert, A. M., Tucker, D. M., & Dishion, T. J. (2017). Dynamic responses in brain networks to social feedback: A dual EEG acquisition study in adolescent couples. Frontiers in Computational Neuroscience, 11, 46. https://doi.org/10.3389/fncom.2017.00046
- Laestadius, L., Bishop, A., Gonzalez, M., Illenčík, D., & Campos-Castillo, C. (2024). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. New Media & Society, 26(10), 5923–5941. https://doi.org/10.1177/14614448221142007
- Laursen, B., Finkelstein, B. D., & Betts, N. T. (2001). A developmental meta-analysis of peer conflict resolution. Developmental Review, 21(4), 423–449. https://doi.org/10.1006/drev.2000.0531
- Li, H., Zhang, R., Lee, Y.-C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. Npj Digital Medicine, 6(1), 236. https://doi.org/10.1038/s41746-023-00979-5
- McBain, R. K., Bozick, R., Diliberti, M., Zhang, L. A., Zhang, F., Burnett, A., Kofner, A., Rader, B., Breslau, J., Stein, B. D., Mehrotra, A., Pines, L. U., Cantor, J., & Yu, H. (2025). Use of generative AI for mental health advice among US adolescents and young adults. JAMA Network Open, 8(11), e2542281. https://doi.org/10.1001/jamanetworkopen.2025.42281
- Meredith, W. J., & Silvers, J. A. (2024). Experience-dependent neurodevelopment of self-regulation in adolescence. Developmental Cognitive Neuroscience, 66(101356), 101356. https://doi.org/10.1016/j.dcn.2024.101356
- Morningstar, M., Grannis, C., Mattson, W. I., & Nelson, E. E. (2019). Associations between adolescents' social re-orientation toward peers over caregivers and neural response to teenage faces. Frontiers in Behavioral Neuroscience, 13, 108. https://doi.org/10.3389/fnbeh.2019.00108
- Neugnot-Cerioli, M., & Laurenty, O. M. (2024). The future of child development in the AI era. Cross-disciplinary perspectives between AI and child development experts. In arXiv [cs.HC]. arXiv. http://arxiv.org/abs/2405.19275
- Phang, J., Lampe, M., Ahmad, L., Agarwal, S., Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., & Maes, P. (2025). Investigating affective use and emotional well-being on ChatGPT. In arXiv [cs.HC]. arXiv. http://arxiv.org/abs/2504.03888
- Powers-O'Brien, A., & Rochman, A. (2025). Generative AI in uncertain times: How teens are navigating a new digital frontier. https://fosi.org/wp-content/uploads/2025/11/Generative-AI-in-Uncertain-Times-FOSI-2025-Research-Report.pdf
- Rapee, R. M., Oar, E. L., Johnco, C. J., Forbes, M. K., Fardouly, J., Magson, N. R., & Richardson, C. E. (2019). Adolescent development and risk for the onset of social-emotional disorders: A review and conceptual model. Behaviour Research and Therapy, 123(103501), 103501. https://doi.org/10.1016/j.brat.2019.103501
- Robb, M.B., & Mann, S. (2025). Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions. San Francisco, CA: Common Sense Media. https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf
- Sahi, R. S., Eisenberger, N. I., & Silvers, J. A. (2023). Peer facilitation of emotion regulation in adolescence. Developmental Cognitive Neuroscience, 62(101262), 101262. https://doi.org/10.1016/j.dcn.2023.101262
- Silverman, M. H., Jedd, K., & Luciana, M. (2015). Neural networks involved in adolescent reward processing: An activation likelihood estimation meta-analysis of functional neuroimaging studies. NeuroImage, 122, 427–439. https://doi.org/10.1016/j.neuroimage.2015.07.083
- Somerville, L. H. (2013). Special issue on the teenage brain: Sensitivity to social evaluation: Sensitivity to social evaluation. Current Directions in Psychological Science, 22(2), 121–127. https://doi.org/10.1177/0963721413476512
- Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J.-H., & Cacioppo, J. T. (2010). Making sense by making sentient: effectance motivation increases anthropomorphism. Journal of Personality and Social Psychology, 99(3), 410–435. https://doi.org/10.1037/a0020240
- Xia, L., Master, S. L., Eckstein, M. K., Baribault, B., Dahl, R. E., Wilbrecht, L., & Collins, A. G. E. (2021). Modeling changes in probabilistic reinforcement learning during adolescence. PLoS Computational Biology, 17(7), e1008524. https://doi.org/10.1371/journal.pcbi.1008524
- Xie, T., Pentina, I., & Hancock, T. (2023). Friend, mentor, lover: does chatbot engagement lead to psychological dependence? Journal of Service Management, 34(4), 806–828. https://doi.org/10.1108/josm-02-2022-0072
- Zhou, T., & Zhang, C. (2024). Examining generative AI user addiction from a C-A-C perspective. Technology in Society, 78(102653), 102653. https://doi.org/10.1016/j.techsoc.2024.102653

Appendix 1
ADVISORY EXPERT CONSULTATION

This report draws on consultations with the following experts. The inclusion of their names does not imply endorsement of the report's conclusions or recommendations; it reflects their participation in interviews or discussions and the contribution of their perspectives to the evidence base.

- David Bickham, Director, Digital Wellness Lab, Boston Children's Hospital
- Emily Cross, Head, Social Brain Sciences Group, ETH Zürich
- Maxime Derian, Research Associate, Luxembourg Centre for Contemporary and Digital History (C²DH), Université du Luxembourg
- Sara M. Grimes, Professor, Department of Communication Studies, McGill University
- Thao Ha, Director, HEART Lab; Associate Professor of Psychology, Arizona State University
- Sameer Hinduja, Co-Director, Cyberbullying Research Center; Professor, Florida Atlantic University
- Kate Blocker, Director of Research and Programs, Children & Screens
- Dan Hipp, Senior Research Coordinator, Children & Screens
- Mimi Ito, Director, Connected Learning Lab, University of California, Irvine
- Melinda Karth, Project Coordinator, Children & Screens
- Pilyoung Kim, Professor of Psychology, University of Denver
- Sonia Livingstone, Professor, Department of Media and Communications, London School of Economics
- Amin Marei, Harvard Graduate School of Education
- Michael Preston, Executive Director, Joan Ganz Cooney Center
- Gregory Renard, Board President, everyone.AI
- Stuart Russell, Professor of Computer Science, University of California, Berkeley
- Sonia Tiwari, Children's Media Researcher
- Ying Xu, Assistant Professor of Education, Harvard University

Appendix 2

Paris Peace Forum, 8th edition 2026

Roundtable

What does AI owe children?

Speakers:

- Clara Chappaz, Former Minister Delegate for Artificial Intelligence and Digital Affairs, France
- Ludwig Charlate. Youth Advocate everyone.AI
- Anne-Sophie Seret, Executive Director, iRAISE
- Mathilde Cerioli, Chief Scientist, Everyone.AI and Lead Researcher, iRAISE Lab
- Lauren Jonas, Policy Lead, Youth Safety and Wellbeing, OpenAI
- Chloé Setter, Product Policy Lead, Children and Teens, Google
- Laurence Devillers, Professor of Artificial Intelligence and Ethics, Sorbonne University
- Irakli Beridze, Head, Centre for Artificial Intelligence and Robotics, United Nations (UNICRI)
- Éric Salobir, Executive Chairman, Human Technology Foundation

https://www.youtube.com/watch?v=Obs-Lo_4kCw

Panel - Forging the Future,
A Dialogue on Beneficial AI for Children, Starting with Principles

Speakers:

- Cécile Aptel, Deputy Director, Global Research and Foresight, UNICEF
- Lauren Jonas, Head of Youth Well-Being, OpenAI
- Michael Preston, Executive Director, Joan Ganz Cooney Center, Sesame Workshop

https://www.youtube.com/watch?v=PATIcrLP_J0

Updated Information about the iRAISE coalition

Updated Information about the iRAISE coalition:
https://www.canva.com/design/DAG2n8nFmKE/mrWHKzbzYfewcEgHH4YIBQ/view?utm_content=DAG2n8nFmKE&utm_campaign=designshare&utm_medium=link2&utm_source=uniquelinks&utlId=h32d396be01

iRAISE

International Research-driven Alliance
for AI Serving Every child

An initiative by