

Children, Young People and AI Chatbots

Mothers' views on the risks of conversational AI, and a shared framework for safeguarding.

Findings from two nationally representative surveys of U.S. mothers — the *AI & Child Safety* study and Wave 1 of the *National Benchmark of U.S. Mothers* — with recommendations for policymakers.

AUTHORS

University College London Kaitlyn Regehr, PhD, Katharine Smales, PhD, and Photini Vrikki, PhD

Count on Mothers Jennifer Brailsford, PhD, and Melissa Lawrence, MPH, MS

A JOINT PUBLICATION BY



Suggested citation: Smales, K., Regehr, K., Brailsford, J., Vrikki, P., & Lawrence, M. (2026). *Children, Young People and AI Chatbots*. White Paper, July 2026.

Abstract

Childhood and adolescence is a critical time for young people’s development, as they learn how to form healthy relationships and develop critical thinking and self-regulation skills. AI Chatbots, designed to foster intimate relationships through personalization and persuasion techniques, create issues to which children and young people are more likely to be vulnerable. This convergence of factors, related to both human development and technological design, heightens the possibility of young people’s exposure to harm: including exposure to misinformation, sexualized and self-harm content, overreliance and deskilling, and the commercial exploitation of their data and privacy. As such, robust safeguarding solutions, designed specifically for children and young people, are needed in this space. The authors recommendation is that until those protections are in place, AI chatbots as they are designed today are not built for children and young people, and their use by minors is best avoided.

Contents

Executive Summary 2

1 Introduction 3

2 AI Chatbots and Children and Young People 4

 Anthropomorphism 4

 Sycophancy 5

 Nudging 5

 Misinformation 6

 Overreliance and desocialization 7

 Data and Privacy 7

3 Current Approaches to Managing Risks relating to AI and Children 8

4 Methodology

10

 Survey 1: AI & Child Safety 10

 Survey 2: National Benchmark of U.S. Mothers (Wave 1) 11

5 Findings

11

Executive Summary

- Research suggests children and young people interact with AI chatbots differently from adults, and specific design features of these chatbots heighten their risk of encountering harm.
- Current US policy does not specifically address the needs and rights of children and young people in relation to AI chatbots. Instead, there is a created expectation by AI companies that parents manage their children's interactions with AI chatbots.
- As a result, this study draws on two large scale nationally representative surveys of U.S. mothers (one of 2,290 and one of 2,818), to gain understanding of their experiences in relation to their children's AI chatbot use and their recommendations for practical and substantial safeguards.
- Mothers reported that they have significant concerns about the design of platforms increasing the risk of harm to children, that AI is frequently introduced outside the home, and that they do not have visibility of their children's activities relating to AI Chatbots, making it difficult for them to act as mediators.
- Among mothers whose children use AI for more than entertainment, a majority report concern in three developmental areas: problem-solving and independent thinking; social skills; and emotional regulation.
- Mothers place less trust in AI companies to act in the best interests of children and families than in any other institution, and that distrust is consistent across the political spectrum, from very conservative to liberal mothers alike. A majority of mothers would support a law requiring stronger safeguards for children using AI.
- The responsibility for creating safe spaces for children and young people must be a shared one between lawmakers, AI chatbot companies, and families.
- Recommendations from mothers include the need for age-appropriate design standards; clear disclosure and transparency; meaningful consent procedures particularly in school settings; limits on data collection; restrictions on manipulative and engagement driven design and finally a recognition that AI platforms have an ongoing duty of care to their users.

1 Introduction

AI is increasingly a part of children and young people's lives,¹ through their use of AI enabled educational tools, social media platforms, search engines, video games, and conversational AI chatbots.² Conversational AI chatbots are now widely used by young people in the USA, with the Pew Research Center reporting that over half of young people use ChatGPT to do their homework³ and Common Sense Media reports that one third use AI companions for social interactions and relationships.⁴

While recognizing the complexity and differences within and between childhood and adolescence, research finds children and young people interact differently with AI than adults⁵ and consequently face particular issues when using AI products which were not designed for them,⁶ including being at greater risk of encountering harmful and inappropriate content.⁷ Although children and young people frequently underestimate the harm conversational AI could pose to them,⁸ research has evidenced that they are concerned by datafication,⁹ others using AI to harm them,¹⁰ mis- and disinformation, and over-reliance¹¹ including forming inappropriate relationships with AI.¹²

Safeguarding solutions are needed for children and young people in this space. This is a crucial time to shape the policy around conversational AI: a technological development which academics, practitioners, and rights organizations agree has the potential to shape children and young people's development and mental health.¹³

¹Common Sense Media (2026a) New Poll Finds Parents Lag Behind Kids on AI and Want Rules and Reliable Information to Help Them, <https://www.commonsensemedia.org/press-releases/new-poll-finds-parents-lag-behind-kids-on-ai> and Common Sense Media (2026b) Talk, Trust and Trade-offs available at https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf

²Neff, G., & Freeman, J. (2026). Written Evidence: Toward Child-Centred AI Safety: Artificial Intelligence Harms & Children's Online Safety. Minderoo Centre for Technology and Democracy. <https://doi.org/10.17863/CAM.129462> and UNICEF (2025) Beyond algorithms: Three signals of changing AI-child interaction, UNICEF <https://www.unicef.org/innocenti/stories/beyond-algorithms-three-signals-changing-ai-child-interaction#:~:text=However%2C%20these%20benefits%20come%20with,allowing%20children%20to%20opt%20out>

³Pew Research Center (2026) How Teens Use and View AI, available at https://www.pewresearch.org/wp-content/uploads/sites/20/2026/02/PI_2026.02.24_Teens-and-AI_REPORT.pdf

⁴Common Sense Media, 2026b.

⁵Yu, et al, (2025) "Exploring Parent-Child Perceptions on Safety in Generative AI: Concerns, Mitigation Strategies, and Design Implications," in 2025 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 2025, pp. 2735-2752, doi: 10.1109/SP61157.2025.00090

⁶Neff and Freeman, 2026.

⁷Mahomed et al (2025) AI, Children's Rights, & Wellbeing: Transnational Frameworks: Mapping 13 Frameworks at the Intersections of Data-Intensive Technologies, Children's Rights, and Wellbeing. The Alan Turing Institute https://www.turing.ac.uk/sites/default/files/2023-12/ai-childrens_rights_wellbeing-transnational_frameworks_report.pdf

⁸Yu et al, 2025.

⁹Wang et al (2025) Child Centred AI, Contextualising Principles and Design in HCI. Forum: Community and Culture <https://dl.acm.org/doi/epdf/10.1145/3757108>

¹⁰Internet Matters, (2025) Me, Myself and AI, <https://www.internetmatters.org/wp-content/uploads/2025/07/Me-Myself-AI-Report.pdf>

¹¹Brandtzaeg, P., Følstad, A. & Skjuve, M (2025) Emerging AI individualism: how young people integrate social AI into everyday life. *Commun. Change* 1, 11 <https://doi.org/10.1007/s44382-025-00011-2> and Internet Matters, 2025.

¹²Internet Matters, 2025.

¹³American Psychological Association (2025) APA urges Senate to regulate AI chatbots, citing grave risks to youth development <https://www.apaservices.org/advocacy/news/chatbots-testimony>; Brewster RCL, Zahedivash A, Tse G, Bourgeois F, Hadland SE (2025) Characteristics and Safety of Consumer Chatbots for Emergent Adolescent Health Concerns. *JAMA Netw Open*;8(10):e2539022. doi: 10.1001/jamanetworkopen.2025.39022. PMID: 41129154; PMCID: PMC12550634; Giovanelli A, Roundfield KD (2025) Adolescent Vulnerability to Consumer Chatbots-Artificial Agents and Genuine Risk. *JAMA Netw Open*;8(10):e2539028. doi: 10.1001/jamanetworkopen.2025.39028. PMID: 41129156. and UNICEF, 2025.

This white paper first qualifies the need for specific risk mitigation for children and young people interacting with AI chatbots, detailing the ways in which children and young people are particularly at risk in relation to harms arising from AI chatbots. It then surveys the current solutions offered to protect children through tools to encourage parental responsibility. It presents the results of two research initiatives which shed light on mothers' firsthand experiences, views, and understanding of their children's AI chatbot usage. It draws out the relationship between mothers' views and the implications for safeguarding children and young people in this space. Finally, it presents mothers' recommendations to policy makers on how to mitigate the unique challenges of children using AI chatbots.

2 AI Chatbots and Children and Young People

The risks from AI chatbots which relate specifically to children and young people evolve from the design features of AI chatbots, coupled with the unique developmental characteristics of minors. The collision of these two factors results in a range of risks.

Anthropomorphism

One of the most commonly cited phenomena in relation to AI chatbots is people's tendency to assign human attributes and form emotional attachments to these technological systems.¹⁴ This phenomenon, termed 'anthropomorphism', has long been observed between humans and machines.¹⁵ This effect is heightened in interactions with AI chatbots because of the personal and emotional data which they collect. This data collection enables a greater level of personalization directed towards the user, which in turn fosters the sense of an intimate relationship.¹⁶ This bond can be strengthened with further data collection and personalization over time.¹⁷ This phenomenon is set to increase over time as AI becomes more agentic,¹⁸ increasingly strategizing, making decisions, and taking action on behalf of users without their direct input, but while using their data to inform these actions.

Young people are particularly prone to forming intense attachments¹⁹ and scholars have found that children and young people have a greater tendency to anthropomorphize AI chatbots.²⁰ This attribution of personality to AI systems by children and young people

¹⁴Placani, A (2024) Anthropomorphism in AI: hype and fallacy. *AI Ethics* 4, 691–698. <https://doi.org/10.1007/s43681-024-00419-4>

¹⁵Weizenbaum (1966) ELIZA—a computer program for the study of natural language communication between man and machine <https://cacm.acm.org/research/eliza-a-computer-program-for-the-study-of-natural-language-communication-between-man-and-machine-2>

¹⁶Bozdağ, A (2025), The AI-mediated intimacy economy: a paradigm shift in digital Interactions, *AI & SOCIETY* 40:2285–2306 <https://doi.org/10.1007/s00146-024-02132-6> and Jean and Esposito (2025) Humans emotionally dependent on AI? Welcome to the attachment economy, <https://blogs.lse.ac.uk/businessreview/2025/08/15/humans-emotionally-dependent-on-ai-welcome-to-the-attachment-economy/>

¹⁷Bozdağ, A, 2025, Jean and Esposito, 2025.

¹⁸Kirk et al, (2025) Why human–AI relationships need socioaffective alignment, *HUMANITIES AND SOCIAL SCIENCES COMMUNICATIONS* | <https://doi.org/10.1057/s41599-025-04532-5>

¹⁹Vasan, N (2025) Why AI companions and young people can make for a dangerous mix, <https://news.stanford.edu/stories/2025/08/ai-companions-chatbots-teens-young-people-risks-dangers-study>

²⁰Kidd, C & Birhane, A, (2023) How AI can distort human beliefs. *Science* 380, 1222–1223. <https://doi.org/10.1126/science.1254444>

means that they form emotionally meaningful relationships with these systems,²¹ and more strongly than adults do.²² These relationships heighten risks, which include mental health, sexual, and representational harms,²³ including self-harm and suicide related conversations and sexualized interactions as well as commercial exploitation through manipulation, persuasion, and targeting.²⁴

Sycophancy

Sycophancy (or insincere flattery) is a design feature which is consistently exhibited by AI chatbots and includes the tendency to sacrifice the truth in favor of pleasing responses to users.²⁵ Research has demonstrated that children and young people are particularly susceptible to peer approval²⁶ and University of Cambridge scholars have shown that AI relationships are more attractive to children and young people because of this sycophancy and the perception of empathy it creates.²⁷

Nudging

Nudging refers to the suggestions and interventions that AI chatbots employ to guide users. They can persuade people to make choices that they may not have made otherwise raising questions about individuals' autonomy and agency.²⁸ Decisions made in this way can have an impact on users' physical or mental health or their financial interests.²⁹ For example, AI chatbots have been shown to be as persuasive as humans,³⁰ a feature which has been attributed to AI chatbots' ability to personalize responses (as a result of data collection), a concept known as 'microtargeting'.³¹ AI chatbots have been shown to be three times as effective at persuading users to buy products than traditional search placement. The same piece of research, out of Princeton, also found that most of the users did not de-

126/science.adi0248; Andries V, Robertson J (2023) 'Alexa Doesn't have that many feelings: children's understanding of AI through interactions with smart speakers in their homes'. arXiv. <https://doi.org/10.48550/arXiv.2305.05597>; Hoffman A, Owen D, Calvert SL (2021) Parent reports of children's parasocial relationships with conversational agents: trusted voices in children's lives. *Human Behav Emerg Technol* 3(4):606–617. <https://doi.org/10.1002/hbe2.271> and Malfacini, K. (2025) The impacts of companion AI on human relationships: risks, benefits, and design considerations. *AI & Soc* 40, 5527–5540. <https://doi.org/10.1007/s00146-025-02318-6>

²¹Namvarpour, M et al, (2025) Understanding Teen Overreliance on AI Companion Chatbots Through Self-Reported Reddit Narratives. *arXiv preprint* <https://arxiv.org/abs/2507.15783> and Yu et al, (2025)

²²Kiuchi, K., Otsu, K., & Hayashi, Y. (2024). Psychological insights into the research and practice of embodied conversational agents, chatbots and social assistive robots: a systematic meta-review. *Behaviour & Information Technology*, 43(15), 3696–3736. <https://doi.org/10.1080/0144929X.2023.2286528>

²³Chang and Razi (2026) A Systematic Literature Review of Generative AI Risks and Harms for Youth, CHI EA '26: Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems Article No.: 54, Pages 1 – 22 <https://doi.org/10.1145/3772363.3798673>

²⁴Jean and Esposito, 2025.

²⁵Sharma et al, (2023) Towards understanding sycophancy in language models [Preprint] arXiv. <https://doi.org/10.48550/arXiv.2310.13548>

²⁶Vasan, 2025.

²⁷Neff and Freeman, 2026.

²⁸Muldoon, R. (2018). Norms, Nudges, and Autonomy. In: Boonin, D. (eds) *The Palgrave Handbook of Philosophy and Public Policy*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-93907-0_18

²⁹Kuyer, P., & Gordijn, B. (2023). Nudge in perspective: A systematic literature review on the ethical issues with nudging. *Rationality and Society*, 35(2), 191–230.

³⁰Höbling, L., Maier, S. & Feuerriegel, S. A meta-analysis of the persuasive power of large language models. *Sci Rep* 15, 43818 (2025). <https://doi.org/10.1038/s41598-025-30783-y>

³¹Salvi, F., Horta Ribeiro, M., Gallotti, R. et al. (2025) On the conversational persuasiveness of GPT-4. *Nat Hum Behav* 9, 1645–1653 <https://doi.org/10.1038/s41562-025-02194-6>

tect that they were being steered towards promotion.³² This design feature increases risk when products which are harmful to people's health are promoted via these methods³³ or when producing deceptive outputs.³⁴ This matters in relation to teenagers, because although young people are increasingly independent and able to morally reason,³⁵ they are also more likely to act impulsively and take risks.³⁶ A tragic example of nudging combined with an intense and intimate relationship with an AI chatbot is evidenced in the case of Adam Raine. Adam began by asking an AI chatbot about his school work and was gradually nudged towards increasing amounts of self-harm content and ultimately to taking his own life.³⁷

Misinformation

The combination of AI chatbot design and children and young people's characteristics provides fertile ground to foster the acceptance of misinformation. One in five young people say they use AI to get news³⁸ but it's well documented that AI chatbots create and disseminate disinformation.³⁹ Disinformation reported by AI chatbots can seem more convincing and believable because interactions with AI chatbots are designed to feel highly personal, private, and akin to a human-to-human relationship.⁴⁰ Adolescents are less likely to question the information they see via AI chatbots,⁴¹ especially when information is given by AI companions who present as friends or mentors.⁴²

³²Salvi et al, (2026) Commercial Persuasion in AI-Mediated Conversations, <https://arxiv.org/abs/2604.04263>

³³Brooks R, Nguyen D, Bhatti A, Allender S, Johnstone M, Lim CP, Backholer K (2022) Use of artificial intelligence to enable dark nudges by transnational food and beverage companies: analysis of company documents. *Public Health Nutr.* 2022 May;25(5):1291-1299. doi: 10.1017/S1368980022000490

³⁴Jones and Bergen (2024) Lies, Damned Lies, and Distributional Language Statistics: Persuasion and Deception with Large Language Models, arXiv:2412.17128

<https://doi.org/10.48550/arXiv.2412.17128>

³⁵Gibbs, J. C. (2003). *Moral development and reality: Beyond the theories of Kohlberg and Hoffman*. SAGE Publications, Inc., <https://doi.org/10.4135/9781452233604>

³⁶Orben, A., Meier, A., Dalgleish, T. et al., (2024) 'Mechanisms linking social media use to adolescent mental health vulnerability', *Nature Reviews Psychology* 3, 407–23 <https://doi.org/10.1038/s44159-024-00307-y>, 5 Rights Foundation (2025) Children and AI Design Code available at <https://5rightsfoundation.com/children-and-ai-code-of-conduct/> and Vasan, 2025.

³⁷Yousif, N (2025) Parents of teenager who took his own life sue OpenAI, *BBC News* <https://www.bbc.co.uk/news/articles/cgerwp7rdlvo>

³⁸Pew Research Center, 2026.

³⁹Spitale, G et al, (2023) AI model GPT-3 (dis)informs us better than humans. *Science Advances* 9, 26 eadh1850. doi:10.1126/sciadv.adh1850 and Glorin et al, (2023) Exploring Ethical Implications of ChatGPT and Other AI Chatbots and Regulation of Disinformation Propagation SSRN: <https://ssrn.com/abstract=4461801>

⁴⁰Neff and Freeman, 2026

⁴¹American Psychological Association (n.d.) Artificial intelligence and adolescent well-being: an APA health advisory. <https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-ai-adolescent-well-being#:~:text=Early%20research%20indicates%20that%20strong,and%20maintain%20real%2Dworld%20relationship.>

⁴²Knoll, L. J., Magis-Weinberg, L., Speekenbrink, M., & Blakemore, S.-J. (2015). Social influence on risk perception during adolescence. *Psychological Science*, 26(5), 583–592. <https://doi.org/10.1177/0956797615569578>; Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological Science*, 22(7), 894–900. <https://doi.org/10.1177/0956797611411057>; van Duijvenvoorde, A. C. K., Peters, S., Braams, B. R., & Crone, E. A. (2016). What motivates adolescents? Neural responses to rewards and their influence on adolescents' risk taking, learning, and cognitive control. *Neuroscience & Biobehavioral Reviews*, 70, 135–147. <https://doi.org/10.1016/j.neubiorev.2016.06.037>

Overreliance and desocialization

Nudging, personalization, and data informed interactions run the risk of children and young people becoming overly reliant on these systems. Professor Mitch Prinstein, of the American Psychological Association, has warned that overreliance on AI could limit young people's ability to build social and emotional skills. Skills which include children and young people's capacity to negotiate difficult conversations, manage conflict, and form trusting real world relationships.⁴³ Some have predicted that this loss of skills will go hand in hand with a decline in peer-based support⁴⁴ and isolation from other humans.⁴⁵ This is particularly acute when combined with teenagers' propensity to feel depressed, to suffer from social anxiety⁴⁶ and to form intense relationships⁴⁷ including parasocial relationships,⁴⁸ all of which could make a relationship with a sycophantic and empathetic AI chatbot more attractive. Indeed, a recent study from Harvard shows that young people between the ages of twelve and twenty-one are turning to AI chatbots for mental health advice at unexpectedly high rates.⁴⁹ Yet, social skills and support networks are essential for young people's future wellbeing and success.⁵⁰ Harms related to overreliance on and deskilling as a result of AI chatbots pose serious threats to young people's cognitive and mental health.⁵¹

Data and Privacy

Research shows that because children and young people are still in the process of forming their capacity for critical thinking and self-regulation, they tend to put their immediate needs above long-term consequences.⁵² As a result, they are at a greater risk of handing over sensitive information to platforms which aim to gather data from them which could be used to profile them.⁵³

A recent report from UNICEF⁵⁴ describes the ability of an AI chatbot, whose responses are personalized according to a child's data, to draw out more information from that child. This then further utilizes the child or young person's data, including to potentially

⁴³APA, 2025 and Youth Endowment Fund, (2025) AI and Teenagers: One in Four Teens Using AI Chatbots for Mental Health Support available at

<https://youthendowmentfund.org.uk/news/one-in-four-teens-turn-to-ai-chatbots-for-mental-health-support-study-finds/>; Wu et al, (2026) The Impact of Chatbots on Adolescent Mental Health Development: A Comprehensive Literature Review. *J Multidiscip Healthc.* 2026 Mar 18;19:579872. doi: 10.2147/JMDH.S579872. PMID: 41873392; PMCID: PMC13005983.

⁴⁴Brandtzaeg, P et al, 2025

⁴⁵Jean and Esposito, 2025.

⁴⁶Orben et al, 2024, 5 Rights Foundation, 2025.

⁴⁷Vasan, 2025

⁴⁸Erickson, S. E., & Dal Cin, S. (2018). Romantic parasocial attachments and the development of romantic scripts, schemas and beliefs among adolescents. *Media Psychology*, 21(1), 111-136. <https://doi.org/10.1080/15213269.2017.1305281>

⁴⁹McBain, Bozick, Diliberti et al (2025) Use of Generative AI for Mental Health Advice Among US Adolescents and Young Adults, *JAMA Netw Open.* 2025;8(11):e2542281. doi:10.1001/jamanetworkopen.2025.42281 and McBain RK, Cantor JH, Breslau J, et al. AI Chatbot Use and Disclosure for Mental Health Among US Adolescents and Young Adults. *JAMA Pediatr.* Published online June 01, 2026. doi:10.1001/jamapediatrics.2026.2015

⁵⁰APA, 2025; UNICEF, 2025.

⁵¹Chalkidis and Sogaard (2026) Brainrot: Deskilling and Addiction are Overlooked AI Risks, FAccT '26, June 25–28, Montreal, QC, Canada.

⁵²5RF, 2025

⁵³5RF, 2025, UNICEF, 2025.

⁵⁴UNICEF, 2025

strengthen the bond between human and machine. This is exacerbated by the proactive nature of AI chatbots, some of which can connect to children and young people's social media accounts or search history to personalize their responses, or take other proactive steps to engage users:

"AI companion apps, those created specifically for deeper relational interaction, don't just wait for user prompts – they initiate interactions, send messages, generate images, and even leave voice notes." (UNICEF, 2025)

It also reports that the majority of children have not provided informed consent for these actions and are unlikely to understand the risks. Even if informed consent is sought from users, Professor Sonia Livingstone's work on children's data and privacy has shown that although children's understanding of data and privacy increases as they get older, there is no 'magic age' of capacity,⁵⁵ a view which is supported by legal scholarship.⁵⁶ This raises questions around how consensual the interactions between AI chatbots and children and young people are.

This overview covers the developmental characteristics of children and young people and demonstrates how they intersect with the design of AI chatbots. This, in turn, highlights a number of risks to children and young people from the formation of unhealthy relationships, the risks of commercial exploitation, exposure to misinformation, the potential for addiction and social and emotional de-skilling as well as threats to their data security and privacy. All this evidences the vital need for safeguards and oversight in this area to protect children and young people.

3 Current Approaches to Managing Risks relating to AI and Children

It is widely agreed that safeguards are needed when it comes to children and AI.⁵⁷ Yet, overall current AI policy has focused on the rights and needs of adults, rather than ensuring that children's rights and needs are met in this space,⁵⁸ failing to recognize that conversational AI can present greater risks to children and young people.⁵⁹ While there are frameworks which recognize child-specific considerations in relation to AI,⁶⁰ much of the guidance globally fails to recognize their unique risks. This includes the White House

⁵⁵Livingstone et al, (2012) Children's data and privacy online. Technology, 58, 2 p 157 – 165.

⁵⁶Do, Y (2025) Beyond Privacy: Regulating ChatGPT for young adults in Educational Contexts, NYU Journal of Intellectual Property and Entertainment Law, 14, 2. <https://jipel.law.nyu.edu/wp-content/uploads/2025/04/JIPEL-Volume-14-Number-2-Do.pdf>

⁵⁷APA, 2025; Brewster RCL, Zahedivash A, Tse G, Bourgeois F, Hadland SE (2025) Characteristics and Safety of Consumer Chatbots for Emergent Adolescent Health Concerns. JAMA Netw Open. ;8(10):e2539022. doi: 10.1001/jamanet-workopen.2025.39022. PMID: 41129154; PMCID: PMC12550634; Giovanelli and Roundfield, 2025; UNICEF, 2025.

⁵⁸5 Rights Foundation, 2025.

⁵⁹Mahomed, 2023.

⁶⁰UNICEF Policy Guidance on AI for children (2021) <https://www.unicef.org/innocenti/media/1341/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf>; UK Information Commissioners Office Age Appropriate Design Code <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/> and the UN-CRC (2021) General comment No. 25 (2021) on children's rights in relation to the digital environment <https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>

memo on the Guidance for the Regulation of AI Applications, which a group of international scholars, including from MIT and the University of Illinois Urbana-Champaign, point out does not make any child specific provisions at all.⁶¹ Currently the U.S. relies on existing laws relating to children online to cover AI chatbots, although there are moves to regulate across individual states.⁶²

One way in which AI companies encourage the mitigation of risks to children is through the provision of parental controls, safety settings, AI activity monitoring tools and AI literacy training for parents available in their family support centers. Meta's AI safety features relating to teen accounts are presented as a way to 'empower parents' to 'make informed decisions', 'block specific AI characters' and 'get insights into topics their teens are chatting about'.⁶³ Likewise, Open AI's parental controls and parent resource page are 'to help families guide how ChatGPT works in their homes'.⁶⁴ In this way, AI companies frame managing the risks related to AI, children and young people as the responsibility of the parents to actively supervise and regulate their children's usage. While companies do not overtly state that parents are responsible for their children's AI interactions, arguably the design of their platform tools for parents creates the expectation that monitoring young people's AI use is within parents' remit.

Currently, research presents mixed findings about how familiar parents are with AI.⁶⁵ Parents do not always understand the types of AI that children engage with⁶⁶ or how they use these different AI applications.⁶⁷ So, while it is recognized that family and parents provide an important learning space for AI Literacy practices where AI skills can be developed together,⁶⁸ it is less clear what parents know about the AI products their children come into contact with and the risks around them, both in and beyond the home. What is clear is that mothers disproportionately take on the role of mediating and negotiating their children's digital media use.⁶⁹

In light of this, Count on Mothers conducted two large-scale surveys into mothers' knowl-

⁶¹Wang et al (2025) Child Centred AI, Contextualising Principles and Design in HCI. Forum: Community and Culture <https://dl.acm.org/doi/epdf/10.1145/3757108>

⁶²Meyer et al, (2025) How Existing Laws Apply to AI Chatbots for Kids and Teens <https://www.law.georgetown.edu/tech-institute/research-insights/insights/how-existing-laws-apply-to-ai-chatbots-for-kids-and-teens-2/>

⁶³Mosseri, A and Wang, A (2025) Empowering Parents, Protecting Teens: Meta's Approach to AI Safety <https://about.fb.com/news/2025/10/teen-ai-safety-approach/>

⁶⁴Open AI (2025) Introducing Parental Controls <https://openai.com/index/introducing-parental-controls/>

⁶⁵J. D. Rubin, E. J. Lombard, K. Chen, R. Divanji, A. Min, E. Dotch, A. Han, K. Dickerson, G. Hayes, and C. Odgers (2025) Navigating AI as a family: Caregivers' perspectives and strategies. White paper. foundry10. <https://www.foundry10.org/resources/navigating-ai-as-a-family> and Abel, C and Magnusson, M (2024) Playdates & Algorithms: Exploring parental awareness and mediation strategies in the age of generative artificial intelligence. <https://ju.diva-portal.org/smash/get/diva2:1869566/FULLTEXT01.pdf>

⁶⁶Wei, X., Chen, X., Yue, L. et al. (2025) Parents' information needs and perceptions of chatbots regarding self medicating their children. *Sci Rep* 15, 30719 <https://doi.org/10.1038/s41598-025-15130-5> and Zhang et al (2025) Exploring families' use and mediation of generative AI: A multi-user perspective. <https://arxiv.org/abs/2504.09004>

⁶⁷Abel and Magnusson, 2024.

⁶⁸Druga, S et al, (2022) Family as a Third Space for AI Literacies: How do children and parents learn about AI together? In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 225, 1–17. <https://doi.org/10.1145/3491102.3502031> and De-laney, V., & Chen, J. J. (2026). Parents' Roles and Strategies in Developing Children's AI Literacy: A Scoping Study. *Journal of Research in Childhood Education*, 40(1), 122–141. <https://doi.org/10.1080/02568543.2025.2581711>.

⁶⁹Pew Research Center (2016) 6 takeaways about how parents monitor their teen's digital activities, <https://www.pewresearch.org/short-reads/2016/01/07/parents-teens-digital-monitoring> and Heaselgrave, F (2025) Unpaid digital care work: Unmasking the parental mediation practices of contemporary mothers. *new media & society* 2025, Vol. 27(1) 281–296 <https://doi.org/10.1177/146144482311744>

edge and understanding of their children's access to and use of AI: the AI & Child Safety study (in partnership with the University of Chicago and University College London) and the National Benchmark of U.S. Mothers (in partnership with University College London). The two surveys sampled 2,290 and 2,818 U.S. mothers respectively, each weighted across region, race/ethnicity, educational level, political ideology, and the age of their children; this work presents a comprehensive and high-quality representation of mothers' attitudes towards AI in the U.S. This paper summarizes the findings of those studies and discusses those findings in relation to the policy insights that they offer.

4 Methodology

Survey 1: AI & Child Safety

This report is based on a nationally representative web-based survey of 2,290 U.S. mothers, *AI & Child Safety: Mothers' Views on a Rising Influence on Kids' Lives* (hereafter the AI & Child Safety report) fielded between October 20 and November 30, 2025. Eligible participants were mothers or primary caregivers aged 18 or older with at least one child under age 21 living at home.

Respondents were recruited through the Count on Mothers national community, partner organizations, and a national online research panel. Quotas and post-survey weighting were applied to align the sample with U.S. benchmarks for mothers by region, race/ethnicity, education level, political ideology, and age of children.

All measures of children's AI use are based on maternal self-report, reflecting how mothers observe, monitor, and interpret their children's exposure across home, school, and social contexts. This approach captures not only patterns of reported use, but also gaps in visibility, oversight, and understanding – factors that are central to parental consent, child safety, and policy-relevant decision-making. A full description of the methodology is provided at the end of this report.

The study was designed to capture mothers' firsthand observations as artificial intelligence becomes increasingly embedded in children's everyday environments. Findings are based on maternal reports of children's AI exposure and use, reflecting how AI is experienced, interpreted, and managed within families.

The research examines several core questions: how and where mothers observe children using AI; how confident mothers feel in understanding AI systems and protecting their children's data; their concerns about AI's impact on behavior, learning, and emotional development; and what safety standards or policies mothers believe are necessary to protect children. By pairing large-scale quantitative data with open-ended qualitative responses, this report offers a nuanced picture of mothers' lived experiences, highlighting gaps in transparency, oversight, and duty of care alongside areas of cautious openness.

Survey 2: National Benchmark of U.S. Mothers (Wave 1)

The second survey is Wave 1 of the National Benchmark of U.S. Mothers, a recurring, nationally representative measure of the conditions shaping American families. Wave 1 was fielded in May 2026 with 2,818 U.S. mothers, each with at least one child under 21 living in the household. The sample was weighted to U.S. population targets for region, race and ethnicity, education level, household income, and mothers' ages. As with the first survey, all measures of children's AI use are based on maternal self-report. The Benchmark tracks five recurring composite measures spanning maternal stress and capacity, family economic pressure, institutional trust, the youth environment and commercial conditions children face, and children's access to wellbeing support; the AI and institutional-trust items reported here sit within those measures. The two surveys are complementary. The first, fielded in late 2025, examined mothers' observations of children's AI exposure and use in depth, while the Benchmark situates AI within the wider set of conditions mothers report and establishes a baseline that future waves will track over time. Because Wave 1 is a single baseline wave, findings describe conditions as mothers report them at one point in time, rather than change over time.

5 Findings

Concern around system design

In the AI & Child Safety report, the most striking finding was that only 6% of mothers feel AI is a positive tool that children should learn how to use, demonstrating that concern is near-universal, with the vast majority of mothers viewing AI as a risk rather than a resource for their children. Mothers' primary concerns were around their children's exposure to unsafe or inappropriate content (63%); confusion between human- and AI-generated interactions (55%) and behavioral or social-emotional changes (51%). The Benchmark points the same way on developmental concern. Among mothers whose children use AI for more than entertainment, 62% agree it makes it harder for their child to learn to solve problems and think independently, 57% agree it makes it harder for them to develop social skills, and 56% agree it makes it harder for them to learn to manage their own feelings. While mothers felt that AI could help their child's schoolwork, they had concerns about how it would affect their development and skills for the future – a concern which is reflected by other parents' views.⁷⁰ Across the concerns found in the AI & Child Safety report, exposure to unsafe content, confusion between AI and human interactions, and social and emotional change; mothers tended to attribute the risk to how AI systems are designed to capture attention, influence behavior, and optimize engagement rather than to children's isolated misuse.

Within the same report, 44% of mothers expressed concerns over data privacy and tracking, which went beyond worries about surveillance or commercialization, rather evidencing an unease around how children's data are used to shape behavior, attention, and beliefs. Concern was most pronounced for practices such as behavioral manipulation, tar-

⁷⁰Rubin et al, 2025.

geted advertising, and undisclosed commercial use of children's data. The Benchmark asked mothers to rate eleven institutions on how much they trust each to act in the best interests of children and families. AI companies drew the least trust of any institution measured: 65% of mothers expressed little or no trust in them, and AI was the only institution where the single largest response was no trust at all (42%).

Data also revealed that mothers' high level of concern around AI spanned political ideology. In the Benchmark, little or no trust in AI companies was a majority view in every political group, ranging from 55% of very conservative mothers to 73% of liberal mothers. Concern was also more prevalent amongst mothers with a higher level of education. This suggests that concern tracks familiarity with how implementing new technologies and systems without appropriate oversight could lead to unintended consequences, rather than an unease about or resistance to digital technology. These concerns are further amplified by limited parental visibility into how and where children encounter AI, contributing to gaps in oversight and accountability.

Institutional Introduction of AI

The AI & Child Safety report found children's primary use of AI was for schoolwork, coming before other aspects like advice and companionship. Forty-three percent of mothers report that their child uses AI for school and one-third of all mothers surveyed report that their child has access to AI through a school-issued device. Breaking this down by age, among mothers of children ages 4–7, over one-quarter report AI exposure via a school device; among children ages 8–12, roughly one-third do so; and by middle and high school, school-based AI exposure is common rather than exceptional. This suggests AI adoption can start at the institutional level where it's compulsory, not optional, imposed by educational settings. Although schools have become an entry point for AI into children's lives there is not an established infrastructure around disclosure, responsibility, or gaining informed consent from parents and their children.

For many families, this institutional adoption changes the nature of consent and oversight. When AI is embedded in required learning environments, parents cannot easily opt out, evaluate alternatives, or independently assess risks.

Lack of Visibility

The introduction of AI systems through schools without clear parental visibility or oversight is also mirrored in domestic spaces. Mothers report that AI is introduced to children and young people through social platforms and search tools. These present contexts which are challenging for families to opt out of, verify how systems operate or engage meaningfully with information and consent procedures around data collection. In the AI & Child Safety report, mothers emphasized that they are not confident about how AI systems use their children's data. Eighty percent of mothers do not feel fully confident they understand how AI companies collect or use their child's data, 11% did not know data were being collected at all, and 28% report not understanding how tracking works.

A lack of visibility in relation to children's AI usage is echoed in academic research. There is evidence that parents/caregivers don't know whether or not their children are using AI tools,⁷¹ or what children are doing in relation to AI chatbots.⁷² While research has shown that a slight majority of parents have had conversations with their children about AI,⁷³ gaps in parental knowledge limit their ability to address the technology's risks in relation to their children.⁷⁴

This lack of transparency is compounded by evidence that parental awareness consistently underestimates adolescent AI use. Mothers' estimates of teen AI use likely understate reality, a gap that becomes clear when compared directly with teen self-reports. In the AI & Child Safety report, mothers report that only about 15% of children ages 13–15 and 18% of children ages 16–18 use AI chatbots or similar conversational AI products. By contrast, recent research finds that over 64% of teens report using a chatbot at least once, with nearly three in ten reporting daily use.⁷⁵

A tendency to underestimate usage is supported in the academic literature. Rubin et al⁷⁶ found that teenagers were more aware of AI companion applications than their parents (50% teenagers knew about virtual friend apps compared to 25% of parents/caregivers). In addition, caregivers often underestimate how often teenagers are using AI products for both educational and social purposes⁷⁷ and are unfamiliar with the ways in which AI is used for media manipulation; chatbot companionship and completing educational assignments.⁷⁸ Moreover, parents' perceptions of the risks presented by AI chatbots are different to the perceptions of young people themselves.⁷⁹ All this suggests that educational and peer based AI engagement occurs more frequently than parents are able to observe, rather than any disengagement on their part.

Notably, the visibility gap persists even among mothers who report actively monitoring and understanding their teen's AI use. Additionally, research has found that although parents support the idea of granular access controls,⁸⁰ which are personalized to their children's ages,⁸¹ parents find it difficult to adapt parental controls in line with the chang-

⁷¹Madden, M., Calvin, A., Hasse, A., & Lenhart, A. (2024). The dawn of the AI era: Teens, parents, and the adoption of generative AI at home and school (pp. 1–56). Common

Sense. <https://www.common sense media.org/research/the-dawn-of-the-ai-era-teens-parents-and-the-adoption-of-generative-ai-at-home-and-school>

⁷²Common Sense Media (2024) The Dawn of the AI Era: Teens, Parents, and the Adoption of Generative AI at Home and School, https://www.common sense media.org/sites/default/files/research/report/2024-the-dawn-of-the-ai-era_final-release-for-web.pdf

⁷³Rubin et al, 2025.

⁷⁴Zhang et al, 2025

⁷⁵Pew Research Center (2025) Teens, Social Media and AI Chatbots 2025 https://www.pewresearch.org/wp-content/uploads/sites/20/2025/12/PI_2025.12.09_Teens-Social-Media-AI_REPORT.pdf

⁷⁶Rubin et al, 2025.

⁷⁷Rubin et al, 2025

⁷⁸Abel and Magnusson, 2024.

⁷⁹Yu et al, 2025 and Han, A et al, (2024). Teachers, Parents, and Students' perspectives on Integrating Generative AI into Elementary Literacy Education. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 678, 1–17. <https://doi.org/10.1145/3613904.3642438>

⁸⁰Atabey et al, (2024) The Second Workshop on Child-Centered AI Design (CCAI) CHI EA '24: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems

Article No.: 492, Pages 1 – 6 <https://doi.org/10.1145/3613905.3636305> and Zhang et al, 2025.

⁸¹Driscoll et al, (2026) Understanding Parents' Desires in Moderating Children's Interactions with GenAI Chatbots through LLM-Generated Probes. In Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26). Association for Computing Machinery, New York, NY, USA, Article 110, 1–33.

ing behaviors of their developing children.⁸²

Across quantitative findings and open-ended responses, mothers repeatedly emphasize that parental vigilance cannot substitute for systemic and ongoing protections. These findings reinforce that parental oversight alone cannot function as the primary line of protection, underscoring the need for built-in safeguards, clear disclosure, and ongoing institutional accountability. Mothers broadly want that kind of protection, with 77% in the Benchmark saying they would support a law requiring stronger safeguards for children using AI and 59% expressing strong support, a level of agreement that held across the political spectrum.

6 Recommendations

Mothers identified six priorities for policymakers in this space, which are accompanied by recommendations for the practical implementation of these points.

RECOMMENDATION 1

Age-appropriate design standards that reflect children's cognitive, emotional, and social development rather than defaulting to adult-oriented systems.

This requires AI companies to recognize the ways in which children interact with AI chatbots across age and stage, including taking into account children and parents' own observations related to this.⁸³ Kurian of the University of Cambridge suggests that long term mechanisms should be put in place to audit AI's impact on children including inappropriate or inaccurate responses by AI chatbots.⁸⁴ Researchers at the University of Oxford argue that design standards should also be implemented which mitigate bias and ensure that systems can be used equally and fairly by children.⁸⁵

RECOMMENDATION 2

Clear disclosure and transparency when children are interacting with AI, including prominent identification of AI-generated versus human-created content and interactions.

<https://doi.org/10.1145/3772318.3791622>

⁸²Maheux and Robb (2026) AI Companions and Youth: What Parents Need to Know, Children and Screens <https://www.childrenandscreens.org/learn-explore/research/ai-social-companions-and-youth-what-parents-need-to-know/>

⁸³Internet Matters, 2025; Neff and Freeman, 2026; Mahomed, 2023 and Feng et al, (2025) The Effectiveness of AI Chatbots in Alleviating Mental Distress and Promoting Health Behaviors Among Adolescents and Young Adults: Systematic Review and Meta-Analysis J Med Internet Res 2025;27:e79850 doi: 10.2196/79850

⁸⁴Kurian, N (2023) AI's Empathy Gap: The Risks of Conversational Artificial Intelligence for Young Children's Well-being and Key Ethical Considerations for Early Childhood Education and Care. Contemporary Issues in Early Childhood, 14639491231206004. <https://doi.org/10.1177/14639491231206004>

⁸⁵Wang, G., Zhao, J., Van Kleek, M. and Shadbolt, N. (2022) Informing Age-Appropriate AI: Examining Principles and Practices of AI for Children. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 536, 1–29. <https://doi.org/10.1145/3491102.3502057>

Those who develop and deploy AI systems should ensure that they disclose information to users clearly and transparently.⁸⁶ Accessible information about AI systems and their design should be provided to parents, including in ways which their children can easily understand. This should include any information about default settings and the parental controls of family accounts.

There should be a clear chain of accountability from the system designers through to the parents or guardians of children and young people which can be used when things go wrong.⁸⁷ This includes the ability for both parents and children to report distressing content or inappropriate responses (Common Sense Media tells us 34% of minors using AI have had uncomfortable responses from AI chatbots).⁸⁸ This reporting should be feasible without having to log into the system and where a human is available to address these concerns.⁸⁹

Transparency includes preventing AI chatbots from claiming professional credentials – such as a therapist,⁹⁰ systems consistently making it clear to children and young people that they are interacting with a machine and directing them towards human advice and companionship.⁹¹

RECOMMENDATION 3

Meaningful consent and parental control, especially in school settings where opting out is often not realistic and families have little visibility into how AI products are actually used.

AI companies, including those whose products are deployed in school settings, must make reasonable efforts to provide a parent with information about how data will be used, kept, and stored in relation to the children. It must obtain the parent's informed consent before the child begins using these systems. Informed consent is not considered valid unless the information is presented in terms which someone can genuinely understand.⁹² If the information is not presented to parents and children in a way they can genuinely understand or in a way which does not clearly outline the implications of the agreement, then consent should not be considered to have been given.

RECOMMENDATION 4

Strict limits on collecting and reusing sensitive data, particularly sensitive personal, behavioral, or emotional information that can be used to shape or manipulate children's behavior.

⁸⁶Jobin et al, (2019) The Global Landscape of Ethics Guidelines Nature Machine Intelligence 1 (9): 389–399. <https://doi.org/10.1038/s42256-019-0088-2>.

⁸⁷Wang et al, 2022.

⁸⁸Common Sense Media, 2026b

⁸⁹5 Rights Foundation, 2025.

⁹⁰Clark A. (2025) The Ability of AI Therapy Bots to Set Limits With Distressed Adolescents: Simulation-Based Comparison Study. JMIR Ment Health.;12:e78414. doi: 10.2196/78414. PMID: 40825182; PMCID: PMC12360667.

⁹¹Kurian, 2023

⁹²Smales K, Lloyd A, Rayner S (2023), "Perry Starlight, Ali Orbit and Kim Cosmos' alien encounter: creating a picturebook as information for children and parents participating in research". Journal of Documentation, Vol. 79 No. 5 pp. 1124–1146, doi: <https://doi.org/10.1108/JD-09-2022-0203>

Children's data should be privy to the highest standards of data protection using data minimization, strong encryption, privacy preserving techniques, and strict non-disclosure to third parties.⁹³ As such, child safety should be prioritized over platform growth incentives, especially with limits on targeted advertising. Platforms should conduct data protection impact assessments (DPIAs).⁹⁴

Platforms should be prohibited from using 'nudging' techniques' to prompt children and young people from revealing more and more information about themselves, including emotional information.

Data collected to comply with a regulatory act or law must be used for that purpose only and held no longer than compliance requires.

RECOMMENDATION 5

Ban manipulative or engagement-driven design. Child-accessible AI systems should not be built to maximize attention, dependency, or behavioral influence.

AI should never cause foreseeable harm including in the form of discrimination, privacy violations, bodily harm, or a loss of trust or skill.⁹⁵ Systems should be designed to minimize the risk of harm to children, including in instances of system failure.⁹⁶

Children and young people's characteristics should not be taken advantage of in relation to platform design.⁹⁷ This includes excluding harmful design patterns which include the promotion of emotional dependency, sycophancy, and anthropomorphism.

Default settings for children and young people should be set at the highest level of control which will disable rewards or incentives, push notifications, and financial transactions. They should also disable nudging responses by the AI system which are not in response to the original prompt.

Crisis detection systems should be made a requirement of AI chatbot platforms, which would detect distress or mental health issues⁹⁸ via patterns of behavior, keywords, or phrases.⁹⁹ Referrals from these systems should be to a human.¹⁰⁰

All safeguarding systems should be thoroughly tested¹⁰¹ and there should be clear penalties for those who violate safety procedures or produce content that is harmful or coercive towards children and young people.

⁹³Wang et al, 2022.

⁹⁴Mahmood, 2023

⁹⁵Jobin et al, 2019

⁹⁶Wang et al, 2022

⁹⁷Wang et al, 2022, 5 Rights Foundation, 2025

⁹⁸McBain, (2025) speaking in Harvard Gazette, How To Regulate AI <https://news.harvard.edu/gazette/story/2025/09/how-to-regulate-artificial-intelligence-ai/>

⁹⁹Zavaleta-Monestel E, Arguedas-Chacón S, Mora-Jiménez J, et al. (March 26, 2026) When Compliance Is Not Safety: The Regulatory Blind Spot in AI Companion Chatbots. Cureus 18(3): e105902. DOI 10.7759/cureus.105902

¹⁰⁰McBain, 2025

¹⁰¹Wang et al, 2022.

RECOMMENDATION 6

Institutional responsibility to ensure that companies and schools---not families alone---are accountable for protecting children from foreseeable risks.

Parents should not be expected to shoulder the responsibility for safeguarding children and young people interacting with AI chatbots. Companies should act with integrity and acknowledge responsibility,¹⁰² making all attempts to foresee the risks to children and young people by design and default.¹⁰³ Special attention and care should be applied to identifying and mitigating potential effects on children's long-term wellbeing and development.¹⁰⁴ Crucially, scholars suggest that compliance needs to be an ongoing process rather than a one-off occurrence,¹⁰⁵ in part in order to account for platforms changing design decisions around monetization or adaptation to different market conditions, where they maximize opportunity potentially to the detriment of users.¹⁰⁶

Ongoing studies should be conducted to understand the impacts of AI chatbots on children and young people so that future risks can be mitigated.

7 Conclusion

This white paper makes clear the risks which AI chatbots present to children and young people. The design of AI chatbot platforms which incorporate anthropomorphism, sycophancy, and nudging towards greater personalization and persuasion techniques, alongside pro-active engagement with users, intersect with the characteristics of children and young people in problematic ways. These characteristics include the tendency to form intense attachments; to attribute personality to non-human objects; to act impulsively; to accept information from those they consider friends; to feel social anxiety and to be developing their capacity for critical thinking and self-regulation. As a result, the collision of AI chatbot design features and the characteristics of minors heighten the possibility of exposure to sexual and representational and mental health harms as well as making them vulnerable to misinformation, a risk of overreliance and deskilling, and the potential to exploit their data and privacy.

Currently, US policy in this space does not specifically address the needs and rights of children and young people in relation to AI chatbots. Instead, there is a created expectation by AI companies that parents manage their children's interactions with AI chatbots. Yet two large-scale nationally representative surveys of mothers, presented in this White Paper, have clearly demonstrated the barriers to this solution. Mothers, who research has evidenced take on the bulk of monitoring their children's digital activities, reported that they have significant concerns about the design of platforms increasing the risk of

¹⁰²Jobin et al, 2019.

¹⁰³5 Rights Foundation, 2025.

¹⁰⁴Wang et al, 2022

¹⁰⁵Zavaleta-Monestel, 2026

¹⁰⁶Soni, S (2026) We want standards, and we want them now—why EdTech needs guardrails, LSE blog available at <https://blogs.lse.ac.uk/medialse/2026/02/27/we-want-standards-and-we-want-them-now-why-edtech-needs-guardrails/>

harm to children, that AI is frequently introduced through educational institutions not the home and that they do not have visibility of their children's activities relating to AI chatbots. This institutional mandate around children and young people's AI usage, combined with the lack of visibility for parents means that the responsibility for creating safe spaces for children and young people must be shared between lawmakers, AI chatbot companies, and families.

Recommendations based on survey data of more than two thousand mothers concludes this report. These recommendations include the need for age appropriate design standards; clear disclosure and transparency when children are interacting with AI; meaningful consent procedures and parental control, particularly in school settings; limits on data collection and secondary use; restrictions on manipulative and engagement driven design; and finally, a duty of no-harm. Together these recommendations call for a shared accountability in relation to children and AI chatbots, rather than one placed solely on mothers.