



Final Statement of the Workshop on Risks and Opportunities of AI for Children

A Common Commitment for Safeguarding Children



Summary

The Pontifical Academy of Sciences, in collaboration with the World Childhood Foundation, chaired by Queen Silvia of Sweden, and the Institute of Anthropology of the Pontifical Gregorian University, convened a high-level workshop to address the opportunities and risks AI presents for children. The conference, held at Casina Pio IV in the Vatican on March 21-22, 2025, included political leaders, AI expert scientists, youth representatives, United Nations and international organizations, civil society, corporate leaders, survivors of abuse, and church officials. While AI has the potential to enhance safety, education, and well-being, it also introduces serious risks.

Key Risks AI Poses to Children:

- AI-Driven problematic behavior & Mental Health Issues: Algorithms maximize engagement, leading to excessive screen time, anxiety, and sleep disorders.
- Privacy Violations & Data Exploitation: AI systems collect children's personal data without stringent safeguards.
- Manipulation & Behavioral Targeting: AI-powered advertising influences children's consumption habits, often without them realizing it.
- Social & Emotional Development Concerns: Over-reliance on AI chatbots and AI-powered toys may limit real-life social interactions.

- Sexual Exploitation & Grooming: AI has been misused to generate child exploitation materials and facilitate online grooming.
- Educational Risks: Excessive digital exposure negatively impacts attentional, linguistic, and cognitive processes, as well as academic performance.

AI's Positive Potential for Children:

- Educational Benefits: AI-driven learning tools can personalize education, expanding access to quality education in particular in underserved regions.
- Protecting from Harmful Content: AI can filter out explicit material and monitor online interactions for safety.
- Early (Mental) Health Detection: AI-powered tools and chatbots can help identify health issues, incl. signs of anxiety or depression in children.
- Enhanced Safety & Surveillance: AI can be used to detect risks in public spaces, schools, and online environments.
- Disaster or Disease Prevention: Rapid response coordination targeted on vulnerable children.
- Promoting Creativity and Innovation: Used responsibly, AI can enhance children's creativity by providing new tools for artistic expression, storytelling, and problem-solving.

A Call to Action for Ethical AI Development: AI must be developed responsibly, prioritizing children's well-being. Stakeholders—including governments, parliamentarians, tech companies, educators, civil society, and faith-based institutions—must work together to establish and enforce ethical regulations, AI governance frameworks, and child-centered policies and interventions. Key recommendations include:

1. Strengthening international governance and regulations to protect children from AI-driven harm.
2. Developing AI systems that prioritize children's safety, privacy, and dignity.
3. Promoting AI's benefits while ensuring transparency, accountability, and fairness.
4. Elevating youth voices in AI governance discussions, and in the design and development of AI systems and products.
5. Global collaboration to ensure that AI serves, rather than endangers, the intellectual potential, dignity and well-being of every child in the digital age.

Risks and Opportunities of AI for Children: A Common Commitment for Safeguarding Children

The Pontifical Academy of Sciences, in collaboration with the World Childhood Foundation,

chaired by Queen Silvia of Sweden, and the Institute of Anthropology of the Pontifical Gregorian University, convened a high-level workshop to address the opportunities and risks AI presents for children. The conference, held at Casina Pio IV in the Vatican on March 21-22, 2025, included political leaders, parliamentarians, AI expert scientists, youth, survivors of abuse, UN and International Organizations, and leaders from civil society and the corporate sector.

Background and Objectives

Artificial Intelligence (AI) is reshaping the digital world at an unprecedented pace, profoundly influencing the lives of children. While AI holds promise in enhancing safety, education, and well-being, it also introduces significant risks. The rapid advancement of AI, accelerated by the widespread digital adoption during the COVID-19 pandemic, has exposed children to increasing online risks, including deep fakes, cyberbullying, and sexual exploitation. AI should serve as a tool to protect, provide, and promote participation, ensuring that children are safeguarded and empowered in the digital world. To address these urgent challenges and opportunities, the Pontifical Academy of Sciences, together with the World Childhood Foundation and Institute of Anthropology of the Pontifical Gregorian University, has convened a high-level workshop with representatives from academia, the technology industry, faith-based institutions, civil society, and youth. We recognize that AI must be developed and deployed with the utmost ethical responsibility to ensure that it protects—rather than endangers—children. AI systems are inseparable from digital platforms, including social media, which shape children’s social, cognitive, and emotional development. It is our moral duty to ensure that these technologies uphold the dignity, rights, cognitive and social-emotional development, and safety of every child.

Emerging Research Insights on AI Risks and

Opportunities for Children

The workshop is based on new scientific insights from various relevant disciplines. Assessing the risks and opportunities of AI requires interdisciplinary cooperation and close interaction between sciences and other stakeholders.

AI Risks for Children

AI can be misused in ways that directly put children’s safety, cognitive potential, and well-being at risk. It has been exploited to generate and disseminate child sexual abuse material, facilitate online grooming, and enable human trafficking. AI-driven surveillance, while often implemented for

safety, can unintentionally infringe upon children's right to privacy, fostering a culture of excessive monitoring that undermines their autonomy and dignity. Additionally, biased AI algorithms may perpetuate existing social inequalities, disproportionately affecting vulnerable children. The increasing use of AI-driven algorithms by social media platforms has raised new concerns about online protection of children. AI algorithms designed to maximize user engagement often prioritize sensational, emotionally charged, or addictive content, which can negatively affect children's mental health and academic performance. The disproportionate impact of AI-driven practices in different groups of children must be recognized in risk patterns, particularly children from the Global South (which account for 75% of the world's children) and those with disabilities—in order to inform targeted and effective policy measures.

1. **Increased Screen Time and AI-Driven-disordered-technology-use:** AI-powered algorithms in digital platforms, including social media, video games, and educational applications, are designed to maximize user engagement. Studies indicate that AI-driven content recommendations may lead to disordered technology use, reducing children's physical activity, sleep and academic performance. Disordered use of technology has been associated with more anxiety and depression, although discussions regarding causality between variables are still a matter of debate. Further, in the research field it is discussed if the data business model aiming at the prolongation of online time takes too much of children's time, which is needed for distinct age-dependent psycho-developmental tasks including sufficient play time and direct social interaction.
2. **AI Chatbots and Emotional Dependency:** The rise of AI-powered chatbots, such as generative AI companions, has introduced new concerns regarding children's emotional development. Over-reliance on AI-driven interactions may hinder children's ability to develop real-life social skills, resulting in lower empathy levels and difficulties with peer relationships.
3. **Collection and Exploitation of Children's Data:** AI-driven platforms collect vast amounts of personal data, including behavioral patterns, location, and preferences. Many AI-powered apps designed for children lack stringent data protection measures, potentially exposing them to targeted advertising, identity theft, and unauthorized surveillance. Often, data are collected without proper consent.
4. **AI-Powered Manipulation and Behavioral Targeting:** AI-driven marketing strategies are increasingly used to influence children's preferences and consumption habits which can lead to commercial exploitation. Studies show that children are particularly vulnerable to AI-powered advertising, as they may struggle to differentiate between organic content and AI-generated advertisements. AI-driven advertising increases children's likelihood of making impulsive purchasing decisions, raising ethical concerns about the manipulation of young minds. AI-generated content can mislead children, spreading fake news or inappropriate material. Deepfakes make it harder for children to distinguish real from fake, right from wrong.

5. **AI-Powered Toys and Their Influence on Child Development:** The growing market for AI-powered toys, such as interactive robots and smart assistants, raises concerns about their impact on children's social and cognitive development. Excessive reliance on AI-based interactions may limit opportunities for real-world socialization, leading to weaker verbal communication skills and less interest in cooperative play compared to peers who interact more frequently with human caregivers and friends.
6. **Potential Reduction in Human Interaction and Learning:** AI-driven social media algorithms personalize and filter content to maximize engagement, often limiting children's exposure to diverse perspectives. This may contribute to the development of "digital echo chambers," where children are only exposed to information that reinforces their existing beliefs. Additionally, excessive digital exposure is negatively linked to attention and academic performance in reading and writing. This could reduce critical thinking skills and hinder their ability to engage with diverse viewpoints in real-world social interactions.
7. **Cyberbullying:** AI systems can potentially promote harmful content, including cyberbullying and exploitation, either directly or indirectly. Since AI is designed to prioritize engagement, harmful interactions such as hate speech or harassment may be amplified. This poses a potential risk to children's safety online, leading to emotional stress and, in some cases, long-term psychological trauma. Moreover, AI-driven platforms frequently promote content that fuels negative comparisons, fostering a distorted sense of self-worth. This is particularly concerning for adolescents, who are still developing their self-identity and emotional resilience. The algorithms can reinforce unrealistic beauty standards, materialism, or toxic behaviors, leading to heightened feelings of inadequacy or anxiety. This problematic behavior with social media also can interfere with sleep patterns, physical activity, and face-to-face social interactions, all of which are vital for a child's healthy development.
8. **Sexual Abuse and Exploitation:** AI has also been exploited by predators to facilitate online sexual abuse and exploitation of children. Through AI-powered chatbots, deepfakes, and virtual worlds, offenders can prey on vulnerable young people. Moreover, AI can be used to generate explicit content, such as deepfake pornography, which can involve the non-consensual use of minors' images. The proliferation of such content increases the risk of exploitation, and the ability of AI to create hyper-realistic images and videos without the need for direct physical contact makes it difficult for authorities to trace and stop offenders. These developments complicate efforts to protect children from online sexual abuse.
9. **Broader AI Risks for Youth:** in addition to child specific risks, broad risks for children shall be pointed out too. These include catastrophic risks from advanced AI systems that could fundamentally affect children's futures, including global security risks or alignment failures in increasingly autonomous systems. AI will also transform economies, potentially creating severe disruption to job markets that today's children will enter.

AI Opportunities for Children

AI also presents opportunities to protect and empower children. Advanced AI systems can detect and prevent online risks in real time, identifying harmful behavior and removing illegal content. In regions with limited access to traditional education and child protection services, particularly in the Global South, AI offers innovative solutions to expand access to learning, health care, and social support.

1. **Protection from Harmful Content:** AI can filter out harmful content and monitor online interactions to protect children from exposure to inappropriate material. For instance, AI algorithms can detect and block explicit content, cyberbullying, or hate speech before children encounter it. Platforms like YouTube have already implemented AI-based tools to remove harmful content, and similar systems can be adapted for other social media platforms. These tools help protect vulnerable children from being exposed to explicit material or violent content that may negatively impact their mental health. In high-income countries with access to advanced technology, these AI systems can be integrated into mainstream platforms to help parents and guardians monitor their children's online activity, giving them greater control over what their children are exposed to. In low-income countries, these tools can be particularly valuable in situations where children may not have parental supervision or access to safe online spaces.

2. **Early Detection of Health Issues:** AI systems can identify early signs of health issues in children and youth. By analyzing patterns in children's social media posts, online behavior, or interactions with virtual assistants, AI can help detect changes in mood or behavior that might signal depression, anxiety, or other mental health concerns. This field is also known as following a digital phenotyping or mobile sensing approach. Such health surveillance systems can then alert parents, teachers, or counselors to intervene early and provide necessary support. In low-income countries, where mental health resources are often scarce, AI could provide a cost-effective way to identify and address mental health issues in children before they become severe. AI-powered apps or chatbots could offer children emotional support, guidance, and coping mechanisms, serving as an initial resource while directing them to professional care if needed. Future AI might provide continuous personalized health monitoring and interventions tailored to each child's unique developmental trajectory.

3. **Enhanced Safety Measures and Surveillance:** In both high and low-income countries, AI can be used to improve child safety in public spaces, schools, and homes. AI-driven surveillance systems can monitor public spaces to identify potential risks or dangerous situations involving children. In schools, AI-based systems can help identify bullying, violence, or unsafe environments, allowing for prompt intervention by school authorities. These systems use facial recognition, behavioral analysis, and predictive modeling to detect and respond to risks. Additionally, in regions with higher levels of poverty or conflict, AI can assist in tracking missing children or identifying children at risk of exploitation. By using AI-powered tools such as facial

recognition, geolocation tracking, and behavioral patterns, authorities can locate and rescue children who are at risk of being trafficked or involved in child labor.

4. Educational Benefits and Access: AI has the potential to transform education for children in both high and low-income countries. In low-income areas, where access to quality education may be limited, AI-powered tools can provide personalized learning experiences, enabling children to learn at their own pace and receive individualized support. AI-driven educational apps can offer interactive lessons, quizzes, and feedback, making learning more engaging and accessible. AI-powered tools assist children with disabilities through speech recognition, translation, and assistive technologies. Moreover, AI can help bridge the digital divide by providing remote learning opportunities to children in underserved areas. In high-income countries, AI can enhance existing educational platforms, creating more adaptive and inclusive learning environments. By tailoring lessons to a child's learning style and progress, AI can help ensure that all children, regardless of income, receive the support they need to succeed. Advanced AI tutors could eventually provide education comparable to having the world's best educators available to every child, potentially transforming educational outcomes globally. AI could transform inclusion for children with disabilities through neural interfaces, advanced prosthetics, or cognitive assistance.

5. Creativity and Innovation: AI can empower children to not only express their creativity but also think critically about real-world challenges and develop innovative solutions. AI-powered tools can allow children to create interactive stories, generate music, or design digital art, expanding their ability to experiment with different forms of expression. Beyond artistic creativity, AI can help children become problem-solvers by enabling them to develop tools that test their knowledge, build interactive learning experiences, or even design AI-driven applications to address issues they care about, such as mental health, environmental sustainability, or social inclusion. For example, children can use AI to analyze environmental data, propose sustainable solutions for their communities, or develop chatbots that provide mental health support to their peers. By engaging with AI as a tool for both creativity and impact, children can cultivate critical thinking, innovation, and a sense of agency in shaping a better future. However, it is essential to guide them in using AI ethically and responsibly, ensuring that technology enhances their imagination rather than limiting independent, critical and democratic thought.

The emerging research highlights both the benefits and risks of AI's growing presence in children's lives. These impacts require a *Child Safety First* ethical approach.

A Call to Action: Ethical AI for Child Protection and Support

Any call to action in the field of AI needs to consider the fast evolving and changing nature of AI. This calls for attention to AI's transformative benefits and risks for children's futures. It is crucial that researchers, regulators, educators, and parents work together to ensure its use does benefit children and minimizes harm, ultimately creating a more secure and supportive digital world for

younger generations. The development and application of AI must be guided by ethical principles that prioritize children's best interests. This requires robust regulatory frameworks, ongoing interdisciplinary dialogue, and responsible AI governance. Workshop participants propose fostering a shared commitment to ethical AI development and deployment, grounded in the following key pillars:

1. Shared Evidence-Based Understanding of the Diverse Impacts of AI on Children

1. Developing a comprehensive, evidence-based understanding of AI's role in children's safety, dignity, and development.
2. Exploring AI's influence across education, governance, child online protection, and industry.
3. Examining AI's impact on youth well-being, including cognitive, emotional, and social development.
4. Identifying both the risks and opportunities of AI for children, ensuring that technology serves as a force for good.

2. Initiatives for Research Partnership and Leadership in AI that Serves Children

5. Building bridges between corporate and public goods related AI research, overcoming constrained access to massive proprietary datasets.
6. Data trusts or cooperatives, with academic-led data commons where individuals or institutions donate or license their data under ethical governance.
7. Academic leadership in creating open, independent benchmarks that even corporate models must be evaluated against, i.e. transparency, fairness, and robustness. This can also include establishing sandboxes for testing and assessing not only the data, models and algorithms but also the regulations, standards and responsibilities.

3. Joint Commitment: Actionable Steps for Child Well-Being

8. While regulations have made significant strides in addressing privacy concerns and limiting harmful content, the challenge remains to protect children from the profound risks posed by AI-driven algorithms designed to engage them. Policymakers worldwide must continue to develop and enforce comprehensive measures that protect children's data, limit their exposure to potentially harmful content, and regulate AI-driven social media platforms when they may foster addictive behavior, and embed children's rights and ethics by design.

9. Given the global nature of AI's risks to Child Welfare, it is essential to enhance international learning about regulations and controls, as well as sharing implementation and enforcement strategies.
10. Outlining concrete measures that stakeholders—governments, tech companies, educators, religious communities, and civil society—can take to enhance child protection, leveraging AI to combat child sexual exploitation, including the detection and removal of abusive material.
11. Promoting AI-driven innovations that expand educational opportunities and improve access to reliable information for children and youth, while ensuring that reliance on such innovations does not disrupt or diminish the development of critical cognitive processes.
12. Fostering multi-stakeholder coalitions and collaborative efforts that promote concrete actions to protect children in the digital age by bringing together governments, science, tech companies, educators, civil society, child advocacy groups, and youth, among other stakeholders, through sustained joint projects and initiatives that advance responsible AI solutions for children. These could establish mechanisms for regular monitoring, evaluation, and accountability in AI deployment, and guide standards and policies.

4. Defined Responsibilities: The Role of Stakeholders

13. Recognizing the unique responsibilities of scientists, the tech industry, policymakers, parliamentarians, civil society, religious communities—including the Catholic Church—, international organizations, and young people themselves as well as parents and families in shaping ethical AI design and use.
14. Addressing the urgent need for rigorous, evidence-based research to inform AI policies that safeguard children's rights, and strengthen related criminal justice systems. Encouraging research specifically on making increasingly autonomous systems safe and aligned with human values.
15. Exploring AI's evolving role in in all relevant systems connected to children for preventing child abuse, advancing education, and supporting digital health initiatives for young people.

6. Ethical and Responsible AI: The Path Forward

16. Advocating for AI policies and regulations that prioritize children's protection in the digital environment, developing sound AI standards and reviewing of related existing legislation.
17. Elevating the voices of young people in discussions on AI governance with digital ethics, and the design of AI systems.

18. Examining the opportunities AI presents for children with disabilities, enabling greater inclusion and accessibility.

19. Promoting AI-driven governance models that serve the best interests of children, ensuring fairness, transparency, and accountability in digital systems. These should adapt to Anticipatory Governance, i.e. governance mechanisms specifically designed to adapt to rapidly evolving capabilities.

As AI continues to shape the digital landscape, we must remain steadfast in our commitment to ensuring that technology serves humanity—especially its most vulnerable members. The Pontifical Academy of Sciences calls upon all stakeholders to work collaboratively, guided by moral responsibility and ethical principles, to harness the power of AI for the good of all children. Only through shared commitment and decisive action can we ensure that AI safeguards the dignity, safety, and well-being of every child in the digital age.