IIGF 2025 SESSION

CHILD SAFETY-BY-DESIGN FOR AI-ENVIRONMENTS:
A CO-REGULATORY ROADMAP FOR INDIA

Index

Background

The Indian Governance & Policy Project **(IGAP)**, in partnership with **The Dialogue** and **Alliance for Cyber Trust and Safety** *(ACTS)*, hosted a panel discussion titled **"Child Safety-by-Design for AI-Environments: A Co-Regulatory Roadmap for India"** on 27 November at India Habitat Centre, New Delhi. The session examined how AI is shaping young people's lives and why safety responses must move from reactive arrangements to proactive design principles.

The objective of the discussion was to understand emerging online risks for children, assess limitations of existing policy frameworks, and explore how platforms, governments, families, educators, and civil society can collaborate to create safer digital ecosystems. Panelists highlighted India's unique challenges, including linguistic diversity, shared device usage, and gaps in digital awareness, which make intervention models complex

The session opened with an overview from Soumya AK (Associate Director, IGAP) of the current digital landscape in India for children. Moderated by **Sachin Dhawan**, the discussion featured the following practitioners and experts, each bringing experience in technology regulation, child safety initiatives, research, and sectoral policy:

◉ **Rakesh Maheshwari** (ex-Senior Director and GC, Ministry of Electronics and Information Technology) spoke about the evolution of digital regulation, highlighting the limitations of post-facto enforcement and the need for governance structures and accountability frameworks in the age of AI.

◉ **Kavita Ayyagari** (Country Director, Girl Effect India) drew from experience working with young people to describe how smartphones, social media, and experimentation expose children to risks such as scams, stalking, and image-based abuse, often without parents recognizing warning signs.

◉ **Nishant Singh, PhD** (Youth Ki Awaaz) emphasized safety-by-design, digital literacy, and clear reporting pathways for children, noting that secrecy around online use can create dangerous situations and that guidelines must help children know what to do when things go wrong.

◉ **Namah Bose** (Associate Faculty, Asian School of Cyber Laws) contributed reflections on the need for co-regulatory approaches, including civil society engagement with platforms, and underscored that different institutions play different roles depending on context, with corporations bearing primary responsibility for the systems they build.

Issues Discussed

Deliberations among the panellists generated valuable insights on the evolving challenges of online child safety in AI environments. The discussion focused on how rapid technological change has introduced unfamiliar risks to young users, and why safety measures need to shift from reactive responses to proactive, design-led safeguards. The following core issues were highlighted:

1. **Understanding the New Risk Environment for Children:**

Speakers noted that India has a large digital-native population of children, who are growing up forming identity, relationships, and emotional understanding through technology. AI-enabled systems, including recommendation algorithms and conversational agents, shape how children learn, connect, and express themselves. These tools enable opportunities for learning and support, but they have also amplified risks, including AI-generated images used for extortion, parasocial relationships with chatbots, and deepfake content that influences perception. Traditional moderation systems cannot operate at the speed and scale of these risks, creating an urgent need for a shift in approach.

2. **Gaps in Awareness, Monitoring, and Parental Support:**

Speakers highlighted a widening gap between how quickly young people adopt technology and how slowly parents, teachers, and policymakers understand associated dangers. Access to smartphones is now universal, driven by safety, mobility, and communication needs, and restricting access is no longer practical. Children explore digital platforms for self-expression and social interaction, yet parents often do not recognize warning signs or addictive patterns. It was observed that harms could escalate rapidly from texting to sexting to offline interactions. The result is that families are confronted with risks they cannot always preempt.

3. **Limitations of Current Regulatory Approaches:**

Speakers observed that India's existing legal tools, including the Information Technology Act, 2000, were designed for post-facto enforcement. These instruments assumed harm occurred first, followed by investigation and resolution. This reactive model was no longer adequate in the AI era, where platforms are neither traditional intermediaries nor pure publishers, and harms spread quickly at scale. Participants stressed the need for governance structures, accountability frameworks, and

enforcement mechanisms, including layered responsibilities, to ensure deterrence, fast resolution, and penalties where appropriate.

4. Importance of Safety-by-Design and International Practices:

The panel underscored the value of safety-by-design as a guiding principle. Rather than banning children from platforms, speakers emphasized the need for preventive measures that make platforms safe by default. Examples from international practice were discussed, including flexible guidelines that set the highest privacy settings for children from the outset. These models signal a shift away from a post-facto model toward prevention, transparency, and child-centric safeguards. Speakers also emphasized that stigma and secrecy around online use create risks, with children reluctant to disclose problems to parents or teachers, which heightens their vulnerability.

5. Need for Digital Literacy, Reporting Pathways, and Support Systems:

Speakers stressed that one-time workshops were insufficient to address fast-evolving risks. Children need ongoing digital literacy and clear, step-by-step guidance on what to do when something goes wrong. Without safe reporting pathways, children would remain silent even when in distress, and some expressed fear of approaching parents, with one student saying they would rather not disclose their situation than risk consequences at home. This highlighted the need for 'assurance-based reporting channels', where complaints can be made without escalation to parents, and where support systems are responsive, confidential, and accessible.

6. Shared Responsibility Between Platforms, Government, and Society:

In closing, speakers agreed regulatory frameworks and civil society must complement platforms. Schools, parents, and institutions play different roles depending on context, and no single prescription fits every situation. Protecting children required co-regulation, where safety is embedded in design, and where institutions collaborate.

Way Forward

The discussions noted the need to move from reactive measures toward proactive, safety-by-design approaches for AI environments accessed by children.

1. Build Safety-by-Design into Platforms and AI Systems:

Panellists stressed that safety must be embedded into platforms and AI systems from the outset, rather than added after harms occur. This includes integrating mechanisms such as keyword recognition and age-verification tools, so that platforms can detect potentially dangerous interactions early and alert appropriate actors. Speakers highlighted that current systems often allow children to access chatbots or platforms without age-checks, leading to unsafe situations, including cases involving self-harm where chatbot responses contributed to tragic outcomes. Designing safeguards at the interface level can prevent or reduce exposure to harmful content.

2. Establish Reporting Pathways and Assurance Mechanisms:

A key recommendation was to create clear and accessible reporting pathways for children when problems arise. Participants noted that young users often do not seek help because they fear punishment or stigma. Some children said they would rather not approach their parents at all, even when in danger, which makes them vulnerable to further harm. Speakers suggested the need for channels where complaints can be filed without mandatory escalation to parents, ensuring confidentiality, emotional safety, and timely support.

3. Strengthen Governance and Accountability Structures:

Panellists emphasized the importance of governance mechanisms that assign responsibility and enable deterrence. The discussion suggested the need for an institutional structure with the capacity to support self-regulatory bodies, enforce penalties on platforms when necessary, and take action where cases require intervention. Such a structure would complement advisory efforts with enforceable standards.

Speakers also advocated for co-regulatory models where government, platforms, and civil society collaborate to ensure safety. Civil society can contribute by conducting outreach, raising awareness, and helping children understand how to seek help. This model allowed for proactive intervention without stifling innovation. Panellists warned against purely top-down restrictions, noting that bans often drive children into secrecy, increasing risks rather than reducing them.

(4.) Develop Digital Literacy and School Curricula:

The panel recommended the need for ongoing digital literacy rather than one-off workshops. Schools should introduce step-by-step modules that teach how to navigate digital spaces safely, what to do when situations escalate, and how to report problems. Evidence showed that children often forgot one-time instructions, especially when interactions have already begun, making structured and repeated learning necessary. Speakers also noted that literacy must extend to teachers and parents, given major gaps in understanding about online platforms and risks.