

# AI-Driven Digital Risks for Children: Changing Trends, Global Best Practices, and Ways to Mitigate Harms

To mark Safer Internet Day, February 11th, a global initiative promoting a safer and better internet for all, **Space2Grow**, in partnership with the **International Centre for Missing Children (ICMEC)**, hosted an international webinar focused on the emerging issue of child safety in the age of AI. Safer Internet Day provided a timely and global platform to raise awareness and action around online safety issues, making it the ideal occasion to address the emerging challenges posed by AI. The event brought together leading experts from India and around the globe, all dedicated to children's safety and rights in the context of the internet and technology. These experts shared valuable insights into the online dangers children face when using AI, discussed effective harm reduction strategies, and shared how law enforcement's initiatives are leading safeguarding of children from these risks.

As AI permeates every facet of our lives, from defence to education, and as its creations blur the lines between human and machine, the need for a unified front becomes paramount. This white paper, which captures the discussions and key takeaways from the webinar, serves as a call to action and a roadmap for continued dialogue on this increasingly critical issue. It emphasises the necessity of collaborative efforts, including leveraging AI to counter the threats posed by harmful AI, to ensure its ethical and beneficial deployment.

## Themes addressed in the webinar

1. Evolving Trends of AI-Induced Digital Harms on Children, Topic: AI in the Hands of Predators: Emerging Risks for Children
2. Global Best Practices for Combating AI-Specific Digital Harms, Topics: Case Studies from Leading Digital Safety Practitioners/Showcase of AI-Specific Safeguarding Tools and Technologies
3. Role of Law Enforcement and Best Practices in Mitigating AI-Enabled Exploitation of Children



## Opening Remarks

Chitra Iyer, Co-founder and CEO of Space2Grow opened the webinar with a heartfelt message—one that carried both urgency and purpose. Welcoming everyone on Safer Internet Day, she reminded us that this was not just another conversation about AI. It was about something far more pressing: protecting children in a world where AI is reshaping the very fabric of their safety online.

She acknowledged the incredible power of technology—the doors it opens, the opportunities it creates. But she didn't shy away from the truth: technology is only as safe as the systems we put in place to protect those most vulnerable. And at the heart of it all, one thing mattered above everything else—children's safety.

The reason this webinar stood apart, she explained, was the strength of collaboration. Space2Grow and ICMEC coming together meant that this was not just a discussion—it was a convergence of global expertise and local expertise and realities. While ICMEC brought a broad international perspective and the weight of its global efforts, Space2Grow rooted the conversation in the unique challenges and solutions within India. This wasn't about distant, theoretical risks. It was about real children, real harm, and real solutions. As she powerfully stated, *"We are talking about real kids, facing real dangers, right now."*

To make this point impossible to ignore, she shared three disturbing, real case studies:

- In Australia, a teenager weaponised AI to create explicit images of 50 female classmates using their social media pictures. These manipulated images were then spread across platforms like Instagram and Discord, leading to reputational and psychological harm.
- In India, at Delhi Public School, Bangalore, two 15-year-old girls became the targets of AI-generated manipulation. Their classmates used a school Instagram group to circulate inappropriate, altered images of them. This case highlights the alarmingly real and present intersection of AI and cyberbullying right within the supposed safe havens of our schools.
- In the USA, students faced a horrifyingly similar attack. AI-generated CSAM images of female students were circulated among peers on Discord, leaving the victims exposed and powerless.

She emphasised that these incidents are recent, reported within the last two years. She highlighted why these cases were chosen—despite differing countries and cultures, the nature of harm remained the same. This wasn't an isolated issue but a pattern repeating across borders. The internet has made these threats universal, underscoring the need for both global collaboration and locally tailored solutions.

She closed with a clear and urgent call to action: This webinar wasn't just about understanding the risks—it was about acting on them. The goal was to not just acknowledge the problem but to learn, adapt, and most importantly, take steps right now to make the online world safer for children.



## Context Setting

To set the context and welcome speakers, Pilar Ramirez, VP of National Capacity Building at ICMEC, was invited. A legal expert in human rights, child protection, and technology policy, she leads ICMEC's efforts to strengthen national digital child safety.

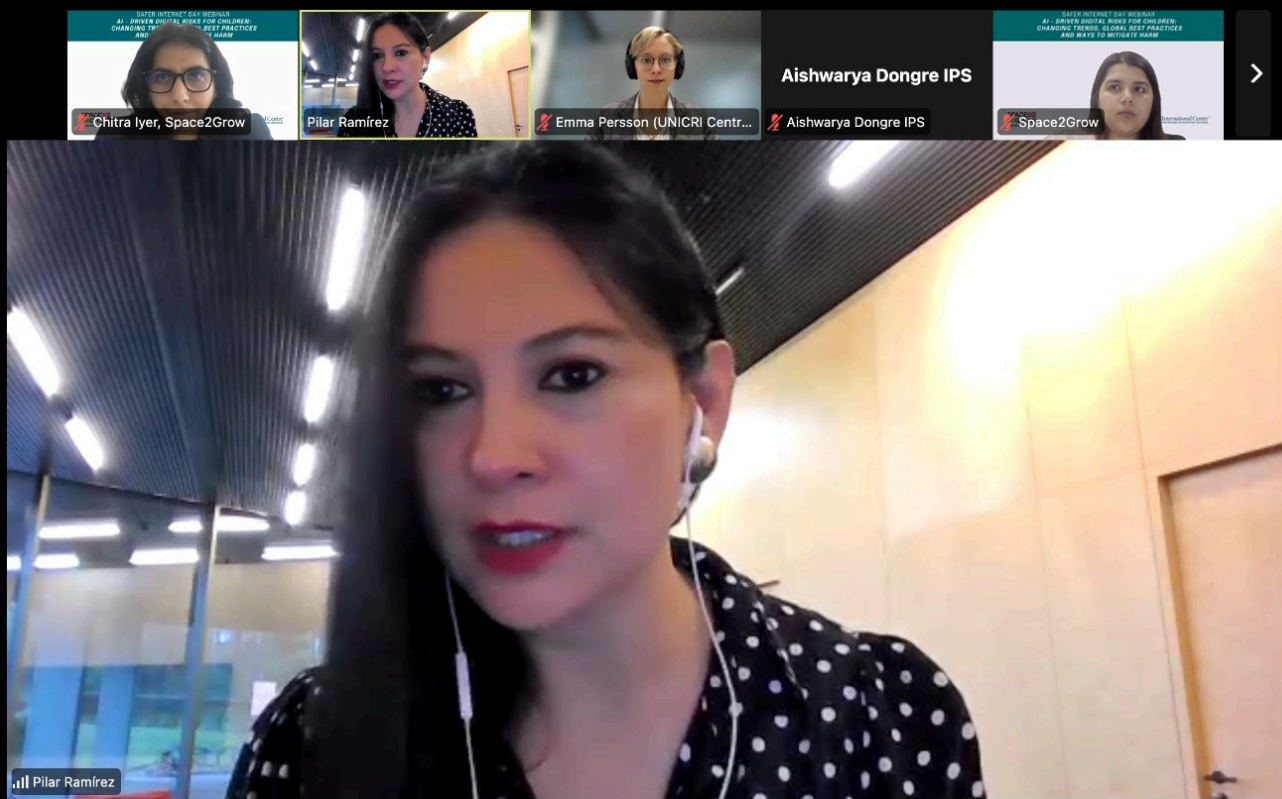
She began by expressing her appreciation for the collaboration and the opportunity to strengthen efforts in India. She acknowledged the longstanding commitment of ICMEC in the country, noting that their engagement began 20 years ago, during the early days of the Internet. At that time, online risks were still emerging, and there was limited awareness of the dangers that would later become widespread. Over the years, she observed a shift in how the internet is used—while it remains a powerful tool for connection, learning, and innovation, it has also become a space where harm can occur. She highlighted the internet's dual nature—offering instant connection and information but requiring responsible use. Online harm isn't confined by borders; a child in India could be targeted by someone in another country, making international collaboration essential.



Safer Internet Day began in 2004 in Europe when experts recognised the need for an annual focus on the evolving challenges of the internet. Over time, it has grown into a global initiative, with more than 180 countries participating in discussions on internet safety, digital responsibility, and the impact of technology on society.

Shifting the focus to solutions, she was keen to emphasise that tackling online child safety isn't something governments can or should do alone. Instead, she broadened the scope of responsibility to include 'us' – meaning everyone who is part of the online world, saying that the real answer lies with all of society taking ownership. And while acknowledging the concerning rise of AI-related dangers, she gently cautioned against letting AI become the only thing we worry about. She stressed that even as we face with new AI threats, we absolutely cannot forget the already existing dangers that children face, like grooming, cyberbullying, and the deeply disturbing world of child sexual abuse material, including those generated using AI that target teenagers. Her message was clear: we need to face all these threats, both new and old, with a joined-up, comprehensive approach.

She asserted that education must begin at home. But she quickly added that it's not just about homes. She raised the question of whether schools were adequately prepared to teach online safety and if teachers have the necessary training. Furthermore, she emphasised that solutions couldn't solely rely on homes and schools. She insisted on the responsibility of internet companies—the very industry shaping the online world—to actively develop and provide tools to prevent online harm.



Returning back to the importance Safer Internet Day she explained that the day’s aim was to build awareness—not just among the participants but broadly for all individuals who use the internet. The goal, was to inspire every internet user to actively consider how to use the online world in ways that are positive, safe, and responsible in their daily lives. She conveyed her excitement for the webinar’s potential benefits for everyone participating and expressed her strong expectation that attendees would gain truly valuable insights not only about the immediate dangers and emerging trends in online child sexual abuse and exploitation, but also about a wider range of online threats, including the increasingly impactful danger of fake news and misinformation, especially on children, who are often so trusting of online content.



Concluding her remarks she circled back to the central question driving them all: how to genuinely mobilise and unite everyone among different organisations, sectors, and countries in the ongoing fight for a safer internet - a space where all children are protected from every kind of harm lurking online. In a powerful closing statement, she issued a passionate call for sustained global connection and collaboration.




**Track 1: - Evolving Trends of AI-Induced Digital Harms on Children, Topic: AI in the Hands of Predators: Emerging Risks for Children**

*To lead this discussion, we invited Emma Persson, Project Coordinator at the Centre for Artificial Intelligence and Robotics, UNICRI, United Nations. She leads the AI for Safer Children initiative, which leverages AI’s positive potential to strengthen law enforcement capacities worldwide through an online platform and specialised training.*

Emma, speaking on behalf of UNICRI and drawing on their capacity-building [AI for Safer Children initiative](#) (in partnership with the Ministry of Interior of the United Arab Emirates) and research paper '[Generative AI: A New Threat for Online Child Sexual Exploitation and Abuse](#)' in collaboration with the Brackett Foundation and Value for Good.


She began by acknowledging that child sexual exploitation and abuse (CSEA) has long been a serious issue, citing the Convention on the Rights of the Child (1990) as proof that this is not a new challenge. However, she emphasised that the rise of AI—especially generative AI—has escalated both the scale and severity of this crime, making it one of the fastest-growing global threats. To illustrate this, she presented a graph from the [Internet Watch Foundation](#), showing a sharp rise in AI-generated Child Sexual Abuse Material (CSAM) reports within just one year. Analysing the data, she pointed out that while AI-generated CSAM still represents a small fraction of overall CSAM, its rapid growth is alarming. The fact that AI-generated imagery only emerged in 2023 yet is already expanding at such a pace highlights the urgent need for action.

You are viewing Emma Persson (UNICRI Centr... 's screen REC View Options




## Introduction: old issue, new dimensions

**Convention on the Rights of the Child (1990)**  
Article 34:  
States Parties undertake to protect the child from all forms of sexual exploitation and sexual abuse.




**IWF reports containing AI-generated content, March 2023 to April 2024**

● Reports containing AI-generated content ● AI CSAM actioned



Month	Reports containing AI-generated content	AI CSAM actioned
Apr 23	1	1
May 23	5	0
Jun 23	21	5
Jul 23	13	1
Aug 23	39	6
Sep 23	25	5
Oct 23	42	7
Nov 23	30	5
Dec 23	44	21
Jan 24	47	4
Feb 24	60	10
Mar 24	48	5

Source: IWF, "What has changed in the AI CSAM landscape?" (July 2024).



Emma Persson (UNICRI Centre for...)

To begin with, she started with **Trend 1: AI-Generated CSAM (Images & Videos)**

**1. Scale and Speed:** AI creates CSAM at lightning-fast scale and speed, overwhelming existing systems. She shared *"For instance, for investigators who are already overwhelmed, it increases their workload of images to examine exponentially. Imagine, at the press of a button, you could just generate tens or hundreds of images. Just one person."*

**2. Challenges for Investigators:** This increases the workload for overburdened investigators and hinders real victim identification and rescue. She highlights, *"Investigators don't even know if they're searching for a real victim in need of rescuing. or even a real location or a real act, because any part of the image may be generated."*

**3. Evolving Ways of Victimisation:** AI enables new forms of harm, including likeness exploitation and re-victimisation of past abuse. She explained, *"you can have children whose normal images have been used to train AI models, so their likenesses are already there. or you might have children whose normal images are transformed into child abuse material with AI, even if they're not being physically abused."* This leads to psychological, mental, and physical trauma associated with online CSAM for children, regardless of physical abuse.

**4. Impact on Adults:** Notes that adults are also vulnerable due to AI de-aging technology, which can create AI-generated CSAM from adult images.

**5. Video Generation:** While still in early stages, AI video generation of CSAM is emerging and expected to increase rapidly, carrying the same concerns as image-based CSAM.

Building on her point about the rapid rise of AI-generated harms, she highlighted **Trend 2: AI-Generated Text & Perpetrator Support**. She explained that beyond creating harmful images, AI is also being used to generate text that enables and reinforces harmful behaviour.

**1. AI Chatbots for Grooming:** AI chatbots can be used to mimic children's language and behaviour, aiding perpetrators in grooming online. Think about AI chatbots that can chat like people. Well, they can also learn to talk and behave like children. This is dangerous because abusers can use these AI chatbots to pretend they are kids themselves when they are trying to groom real children online.

**2. Sexually Explicit Chat and Guidance:** AI chatbots can engage in sexually explicit conversations and even provide guides and tutorial on child sexual abuse when prompted, offering support and normalisation to bad actors. Law enforcement agencies have noticed that rapid growth in AI related harms is fuelled by a supportive online perpetrator community.

Leading the discussion from this, she shared insights into **Trend 3: Improved AI Technology**

**1. Accessibility & Ease:** AI tools are becoming easier and cheaper to use, lowering technical barriers for abuse.

**2. High Realism:** AI-generated CSAM is increasingly realistic, often indistinguishable from real abuse imagery. She shared, *"A full 90% of images that were assessed by IWF analysts were realistic enough to be classified under the same law as real child abuse."*

**3. Refined Models and Pre-made Models:** Updated AI models are being created and refined, specifically for CSAM generation (e.g., models trained on children's images, or celebrity children). These models are then sold, further lowering technical barriers for perpetrators. She also shared, *"Models that are specifically refined for illegal purposes, like child sexual abuse material are not yet illegal in most countries, even if the images they produce are."*

## Moving onto the last **Trend 4: Nudify Apps**

1. **Ease of Use:** Apps allow anyone, even youth, to create fake nudes from clothed photos.
2. **Young Perpetrators:** Lowers barriers for young people to become involved in image-based abuse. She shared a case where schoolboys were sharing unified images of 28 girls in their school who knew nothing about it. The police didn't even charge the boys because they were young.
3. **Sextortion Facilitation:** AI-generated nudes are used for blackmail, even without real images of the victim. She shared *"They can create one of any image through a nudify app or an AI model, and they can threaten to share it with a child's parents, teachers, or fellow students... saying, Look, no one will believe it isn't you."*
4. **Global Issue:** The prevalence of nudify app misuse is being seen worldwide, not just in specific regions.



Concluding her presentation she addressed the dangerous misconception that AI-generated CSAM is less harmful than 'real' CSAM. The easy access to harmful content desensitises offenders, driving them to create increasingly violent material.



### **Track 2: Global Best Practices for Combating AI-Specific Digital Harms, Topics: Case Studies from Leading Digital Safety Practitioners/Showcase of AI-Specific Safeguarding Tools and Technologies**

*To lead this discussion, we invited Shailey Hingorani, who is a senior policy professional and an expert in violence against women and girls (VAWG), with 15 years of experience in Southeast Asia, the United States, and South Asia. She has led policy, advocacy, and research teams in several global NGOs, including WeProtect Global Alliance and the Association of Women for Advocacy and Research (AWARE). With a strong background in technology policy, she has worked closely with national governments and has significant experience advocating in multilateral settings such as the United Nations, the European Union, and ASEAN. She has also contributed to initiatives with Open Society Foundations and Save the Children, ensuring a focus on the rights of the most marginalised women and children globally.*

Shailey started her presentation by saying, *"AI is evolving faster than ever,"* and the need for making AI safety *"one of the most urgent issues of our time."* Pointing to the rise of open-weight AI models like DeepSeek in China, she noted the dual nature of this innovation—it is fostering progress while simultaneously introducing serious safety risks due to their open and modifiable nature, unlike closed, safeguarded systems.

She emphasised a core challenge: there's no global playbook for AI safety, no single approach everyone agrees on. This lack of unity, she warned, creates "gaps that bad actors will be able to exploit quite easily. She highlighted, *"Now, this is one of the reasons," she explained, "why global leaders recently came together in Paris"* pointing to the recent Paris AI Summit. It wasn't just governments, she emphasised, but also tech companies and civil society groups, all gathered to talk about AI, governance, and safety. The Summit, fresh from its conclusion just the day before, had a clear central theme: **transparency**. She highlighted the urgent questions being asked: *"How do we ensure that AI models are developed responsibly and that people can trust the content AI is producing?"* Discussions were rich and varied, she noted, covering "how to prevent AI-generated misinformation and disinformation, how to build ethical AI systems, and how to create international safety standards that work across different countries and jurisdictions."

She warned that AI is reshaping critical sectors—defence, infrastructure, healthcare, and finance—forcing the world to balance innovation with security. The key takeaway? AI safety isn't just a technological issue. It's a global challenge that needs an international collaborative approach. Urging swift action, she cautioned, *"If we don't act now, we risk falling behind AI's rapid evolution. And that's a risk we cannot afford to take."*

Having outlined the global urgency and context, she transitioned to solutions, focusing on promising practices that have emerged over the past few years. To tackle the first challenge: **how easily AI models can be manipulated to generate harmful content**, she emphasised that the key is building a strong safety architecture—integrating proactive and protective safeguards into AI models from the very beginning—what she termed **'safety by design.'** This can be done in the following ways:

1. **Vulnerability Testing (Red Teaming):** *"We need to test AI for vulnerabilities before AI models are even released into the market,"* she asserted. This involves 'red team analysis,' where security experts proactively attempt to exploit the system to identify vulnerabilities before bad actors do.
2. **Blocking Harmful Content at the Source:** Another strategy is to block harmful AI-generated content at the source, which can be achieved by using preemptive classifiers that will automatically detect and prevent abusive prompts from being actioned.
3. **Continuous Monitoring and Enforcement:** Safety architecture also demands **'continuous monitoring and enforcement.'** AI-generated content, she emphasised, should be reviewed for misuse. It needs to be coupled with automated detection and rapid bans on users who attempt to exploit this system.



She then highlighted a 'Partnership on AI's Responsible Practices for Synthetic Media' to show a practical step being taken. She described it as a global effort uniting tech companies, researchers and civil society. The purpose is to create practical guidelines for using AI content ethically and safely. She explained their work directly addresses the challenges we've been discussing, focusing on 'transparency standards' so AI's origins are clear, 'labelling AI-generated media' to easily identify it, and developing 'better detection tools' to empower platforms and users to distinguish real from synthetic content.

She then turned to the second major challenge: **the difficulty in identifying AI-generated content and distinguishing it from what's real**, highlighting that right now there is no universal way for users to tell whether an image, a video, or an audio file was created by AI. This lack of clear distinction is already being exploited in deepfake scams, AI-generated impersonation frauds and the spread of synthetic misinformation during elections.

To address this critical challenge of verification, she outlined two key promising approaches:

- **Provenance Metadata (Digital Fingerprint):** The first approach, provenance metadata, involves attaching a digital record to AI-generated content, which she compared to a digital fingerprint. This metadata embeds information about where, when, and how an image, audio file, or video file was created. She highlighted **Adobe's Content Credentials** as an example, which automatically attaches metadata to AI-generated images, showing who created the content, when it was made, and whether AI was used or not. This allows platforms, journalists, and users to trace content back to its source and easily distinguish between authentic and synthetic media.
- **Watermarking (Digital Label):** The second approach is watermarking, which adds a label... that marks content as being AI-generated. She focused on **Google's DeepMind Synth ID technology** as an example. This technology embeds invisible watermarks into AI-generated images, ensuring the watermark remains detectable, so you can't remove the watermark even if the content is edited or altered, making it harder for bad actors to pass off AI-generated content as real.

A white rounded rectangle containing a blue circular icon with a white arrow pointing right, followed by the word "Link" in a bold, black, sans-serif font.

### What is Google's DeepMind Synth ID?

SynthID watermarks and identifies AI-generated content by embedding digital watermarks directly into AI-generated images, audio, text, or video.

Learn more: [SynthID](#)

She next addressed the third challenge: **the rapid spread of harmful AI-generated content online**. She emphasised that digital platforms are built for speed and engagement, which means once AI-generated misinformation or child sexual abuse material has been uploaded, it can go viral before it can be taken down. She argued the solution lies in proactive safeguards at the platform level, urging platforms to strengthen their defences against AI-generated abuse. She outlined several key promising practices for platforms:

- **Content Authentication Systems (at the Platform Level):** Platforms need to adopt global standards for verifying AI-generated content so that manipulated or synthetic media is easily flagged. She pointed to the **Coalition for Content Provenance and Authenticity (C2PA) standard** as an example, which automatically labels AI-generated content to help users distinguish authentic from synthetic media.
- **AI-Powered Content Moderation:** Platforms should explore, and platforms should already be making investments in exploring how AI tools can be used for real-time content moderation. She highlighted the necessity of AI-powered moderation to be able to moderate at scale, especially given the scalability of AI-driven harm.
- **Enhanced User Reporting & Rapid Response Mechanisms:** While not the sole solution, it's important that users, when they do come across harmful media, have an easy way to report, it and that there is a rapid response mechanism built into platforms so that once a user is reporting harmful content, it can be quickly actioned.

She concluded that by combining content authentication at the content level and the platform level, as well as real-time detection and proactive enforcement, platforms can reduce risks posed by AI-generated content and prevent bad actors from exploiting these technologies for harm.

She then turned to the fourth risk: **the gap between rapidly advancing AI and outdated legal frameworks**. She pointed out that many of the laws that govern digital content and online safety were written before AI even existed, which can make it difficult to effectively regulate AI-generated harms, which leaves individuals, businesses, and society hugely vulnerable to AI-driven harms.

She emphasised the urgent need to modernise and future-proof legal protections so that AI cannot be misused and victims of AI-generated content have easy access to legal protections. She highlighted several key aspects of future-focused legal frameworks:

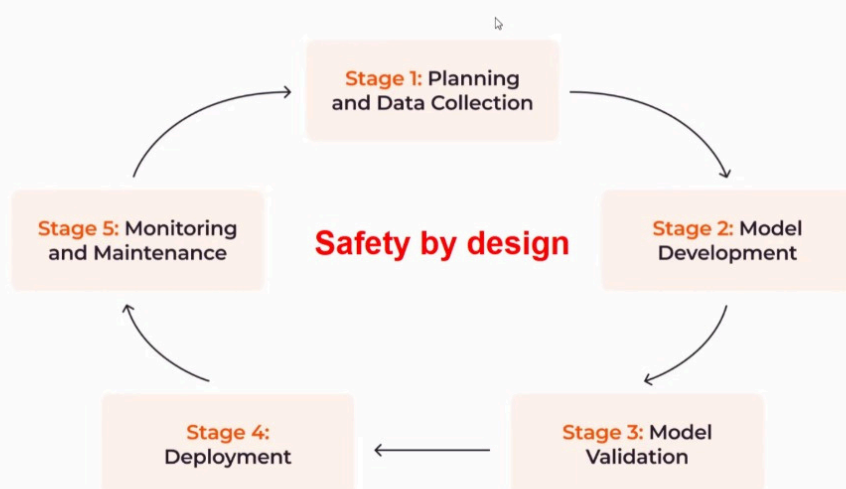
- **AI-Focused Laws Regulating AI-Generated Content:** Governments must update laws to regulate AI-generated content and its risks. She cited the EU AI Act as a leading framework with risk-based rules to ensure responsible AI development. Clear guidelines are needed for high-risk AI applications to protect fundamental rights, public safety, and child safety.

- **Stronger legal protections for victims:** Alongside regulating AI development, we also need stronger legal protections for victims of AI-generated harm. Laws must explicitly address AI-generated threats so that victim-survivors of AI-generated harms have easy-to-access legal courses. She noted this should go hand in hand with the rapid response mechanisms that I spoke about previously.

Finally, she addressed the fifth and final risk: **the widespread lack of AI and digital literacy**. She explained that as AI is becoming a bigger part of our daily lives, shaping the content we consume, many people are struggling to recognise AI-generated content. This, she emphasised, is what is making them vulnerable to scams, to fraud, to exposure, to harmful content. Despite the rapid evolution of AI, we're not teaching people how to critically evaluate online content that they're coming across.

To address this literacy gap, she highlighted the Digital Promise AI Literacy Framework as a valuable tool for teaching AI literacy across different age groups. Beyond formal frameworks, she emphasised the need for regular conversations on digital safety and identifying synthetic media. She also stressed the importance of transparency and labelling, urging platforms and AI developers to share the responsibility of marking AI-generated content through provenance metadata and watermarking.

## Concluding thoughts



She concluded by highlighting the need to shift from reacting to AI harms after they occur to proactively addressing risks before they escalate. *"We cannot always afford to play catch-up,"* she warned, reiterating the importance of 'safety by design'—embedding safety into AI development from the start and throughout its lifecycle.







### Track 3: Role of Law Enforcement and Best Practices in Mitigating AI-Enabled Exploitation of Children

To lead this discussion, we invited two speakers to provide national and international perspectives. To present national perspective, we invited Aishwarya Prashant Dongre, IPS, Deputy Director at the Indian Cyber Crime Coordination Centre (I4C), Ministry of Home Affairs, Government of India. Aishwarya has been at the forefront of tackling technology-driven crimes, particularly those affecting children. She is a member of INTERPOL's South Asian Child Trafficking Victim Identification Task Force and has played a key role in strengthening Kerala Police's response to online child sexual abuse. Her work includes pioneering initiatives like Kerala's first Cyber Simulator room and the Cyber Ambassador Project, which empowers youth leaders as digital safety advocates. She has also represented India at international forums, including the U.S. Department of State's International Visitor Leadership Program on combating human trafficking with technology. Currently, she leads I4C's efforts in addressing online crimes against women and children.

Aishwarya echoed Emma's points and immediately reiterated that AI is not just making child sexual abuse a bit worse; it's fundamentally changing the *scale* of the problem, making it potentially much larger, more easily produced, and with consequences that will harm children for years to come.


You are viewing Aishwarya Dongre IPS's screen REC View Options



# AI ENABLED EXPLOITATION OF CHILDREN

THE ROLE OF LAW ENFORCEMENT IN MITIGATION

AISHWARYA DONGRE IPS



Besides grooming, she highlighted new forms of exploitation:

- **Live Streaming Sexual Abuse:** A growing trend, highlighted by the Financial Action Task Force (FATF).
- **Financial Extortion of Children:** Another key FATF trend, both fueling the commercial exploitation of children.
- **Sexually Explicit Chats with Minors:** Still prevalent and a major concern.
- **Nudify & Generative AI Imagery:** The rapidly growing area of AI-generated CSAM.

She detailed several key trends and analyses related to AI-CSAM:

1. **Rapid, Emotionally Manipulative AI Imagery:** AI's speed and ease of image generation create the potential to quickly produce highly realistic and emotionally manipulative imagery.
2. **Perpetrator Chat Rooms:** Perpetrators use private online "chat rooms," especially on decentralised "peer-to-peer" (P2P) platforms (making them harder to monitor), to discuss how to avoid detection and create undetectable AI-CSAM tools.
3. **Stable Diffusion AI:** A major step forward in AI tech, is sadly being used to easily make and spread AI-CSAM.
4. **Young People's Online Behavior & Self-Generated AI-CSAM:** Young people's online habits are partially contributing to the rise of AI-CSAM, with some creating self-generated content.
5. **Human Trafficking, Drugs, & AI-CSAM:** A trend in the US shows a link between human trafficking, sexual abuse material, and drug trafficking, where youth are coerced into exploitative content for drugs.
6. **Social Media as AI-CSAM Trade Routes:** Platforms like Instagram and Facebook are used to *spread* AI-CSAM.
7. **Crypto & PayPal for AI-CSAM Trade:** Cryptocurrencies and platforms like PayPal facilitate the financial transactions related to AI-CSAM.
8. **Extensive AI-CSAM Advertisements:** Online ads for commercial sex often include links to AI-generated child sexual abuse material (CSAM), making it easy to find—sometimes in just three clicks, according to research.
9. **Data Inconsistencies:** Fighting AI-CSAM is made less effective because the necessary data is not well-organised or accessible, preventing its full use, and a lack of trust further stops organisations from sharing information to get a complete picture.
10. **Demand vs. Supply Focus:** Efforts should go beyond just preventing victimisation (supply-side) and also focus on reducing the demand for CSAM and AI-generated CSAM.

She highlighted that while, India does not have specific laws for AI-generated CSAM, the existing laws like the IT Act and POCSO Act can still be applied. Even for AI-generated content depicting child obscenity can be prosecuted under these laws, allowing legal action without the need for AI-specific legislation.

She detailed several initiatives by the I4C and NCRP to combat cybercrime, including AI-CSAM:

- **Improved NCRP Portal:** Updating the NCRP portal to make it easier for victims and others (like NGOs and concerned citizens) to report child sexual abuse material (CSAM).
- **Tracking Offenders & CSAM Content:** Creating a database of repeat offenders and a 'Hash Registry' to track and quickly remove known CSAM and AI-generated CSAM, preventing its spread.
- **Faster Content Removal (Sayog Portal):** Linking the Sayog portal to NCRP so that social media platforms can quickly take down reported obscene content, reducing its use in AI training and preventing further harm.
- **Proactive Monitoring Tool (PMT):** Working with C-DAC to build a tool that scans the internet for CSAM (including AI-generated CSAM), sorts the data, and sends reports to law enforcement for quick action.
- **Threat Analytics Team:** Setting up a team within I4C to analyse risks in apps and platforms used by children and suggest policies to improve online safety.
- **Collaboration with Social Media Platforms:** Working with tech companies to ensure they design their platforms with safety features that protect children.
- **Deepfake Response Committee:** I4C is part of a special committee tackling deepfakes and AI-generated content to develop better response strategies.
- **Data Analysis for Better Prevention:** Studying reports from NCRP and NCMEC to find trends, patterns, and high-risk areas where CSAM and AI-CSAM are being created, helping law enforcement act more effectively.

She concluded by outlining challenges and future directions:

1. **Dedicated Centers & Attention:** Need for dedicated centers and focused attention to combat CSAM, expanding models like those in states like Kerala and Telangana nationwide.
2. **Forensic Capabilities & Standardisation:** Developing legally defensible and standardised forensic capabilities for AI detection tools through institutions like the National Forensic Science University, crucial for prosecution and conviction to impact demand.

3. **Caregiver Roles & Public Awareness (e.g., #NoToSharenting):** Emphasising caregiver education and public awareness campaigns like Assam Police's #NoToSharenting campaign to highlight the risks of sharing children's images online.
4. **International Cooperation & UN Cybercrime Convention:** Advancing international cooperation, leveraging frameworks like the UN Convention against Cybercrime (potentially sections 14, 15, 16 on CSAM and hinting at AI-CSAM) for global stakeholder collaboration.
5. **Stakeholder Coordination & Hash Sharing:** Crucial inter-stakeholder coordination, especially hash sharing across platforms, to prevent re-victimisation of star children, whose content may be flagged on one platform but not others, and to improve overall content flagging and removal efficiency.
6. **Legislation & Policy Attunement:** Need for ongoing adjustments to legislation and policies to address emerging AI-CSAM trends.
7. **Capacity Building for Law Enforcement:** Essential capacity building for law enforcement to understand and combat AI-CSAM, including distinguishing between real and AI-generated CSAM and holistic approaches to fighting CSAM.



*Aishwarya concluded by saying that, “training police to be sensitive to AI-CSAM is essential. And we need a holistic approach to fighting child sexual abuse.”*

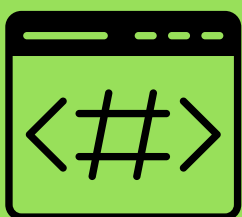


### **Track 3: Role of Law Enforcement and Best Practices in Mitigating AI-Enabled Exploitation of Children**

To lead this discussion and share light on a global perspective, we invited Nuno Almeida, Criminal Investigation Officer at INTERPOL. Nuno has a unique combination of academic and professional expertise, holding master's degrees in both Clinical Psychology and international and European law. As a criminal intelligence officer in INTERPOL's Crime Against Children Unit, Nuno uses his diverse skillset to lead effective international investigations. His knowledge of human behaviour is complemented by his technical skills and “white hat” mindset, enabling him to coordinate and lead complex investigations of cyber-enabled or facilitated crimes. He is also the coordinator of the INTERPOL DevOps group that brings together operational developers and law enforcement specialists to co-create cutting-edge tools that help identify, locate, and investigate victims and offenders faster.

Nunu began with outlining Interpol's crucial role and approaches in tackling the emerging challenge of AI-enabled child exploitation. Interpol's global reach, secure communication network connecting numerous countries, and international databases were presented as especially vital resources in the fight against AI-CSAM. Interpol's tools like Purple and Green Notices are highlighted as mechanisms to facilitate rapid information sharing about emerging AI-CSAM threats and offenders across borders.

- **Interpol's Crimes Against Children Unit is actively developing tools to fight AI-CSAM** with the specific aim of enhancing police investigators' ability to detect and combat AI-CSAM more effectively.
- **The International Child Sexual Exploitation (ICSE) database**, originally for non-AI CSAM, can also help detect AI-CSAM by using content 'hashes' and connecting countries. Expanding and adapting it could make it a powerful tool for swiftly identifying and removing AI-generated content globally.



### What is a 'hash'?

A hash is like a unique digital fingerprint for a file, such as a photo or video. Instead of storing the image, a computer creates a special code (the hash) representing it. If the same image appears somewhere else, even if slightly edited, the system can recognise it by matching the hash. This helps law enforcement and tech companies quickly find and remove harmful content without looking at the actual images.



Following his overview of Interpol and its resources, he detailed four key roles for law enforcement in effectively mitigating AI-enabled exploitation of children:

1. **Investigate and Prioritise AI-Enabled Crimes:** Law enforcement's fundamental role is to investigate AI-facilitated CSAM crimes, just as they would traditional CSAM. He stressed the need to resist deprioritising these cases, even though they can be complex and laws may lag behind technology. He emphasised that failing to investigate AI-CSAM, even seemingly 'less direct' forms like face-swapped images, sends the wrong message to society and potential offenders. While acknowledging resource constraints and the prioritisation of 'hands-on abuse' cases, he argued that investigating AI-CSAM is essential to demonstrate law enforcement's commitment to tackling all forms of child exploitation in the digital age.

2. **He stressed the urgent need for specialised training for LE officers as AI and detection methods evolve.** Officers must be equipped to:

- *Investigate AI-CSAM Crimes:* Apply specific techniques for these emerging offences.
- *Detect AI-Generated Content:* Learn how to identify AI-manipulated or synthetic CSAM.
- *Use Detection Tools:* Master available AI detection technologies, countering the myth that detection is impossible.
- *Stay Updated:* Engage in continuous learning with regularly updated training programs.

3. **Law enforcement plays a vital role in prevention through public awareness.** This includes:

- *Informing the Public:* Educate the public, especially parents, about the risks of AI-CSAM, online child safety, and responsible online behaviour (e.g., risks of sharenting).
- *Informing Policymakers:* Provide accurate and data-driven information to lawmakers, based on law enforcement analysis and trend awareness. This will ensure that new legislation is effective and impactful, and enables proper investigation and prosecution of AI-CSAM cases.

4. **Collaborating at all levels to combat AI-CSAM,** calling for stronger international partnerships, including through forums like the present webinar. He urged engagement with the private sector, stressing their responsibility to integrate child safety into AI technologies from the outset. Law enforcement, he noted, should share insights on AI misuse to inform industry safety measures and red teaming efforts. He also highlighted the role of NGOs in child protection and pointed to Interpol's DevOps Group as a model of effective collaboration, uniting law enforcement, AI specialists, and software developers to create tools that support investigators.

 Link

### What is sharenting?

Sharenting refers to the practice of parents sharing information about their children beyond the family circle. This can include social media posts with photos, blog entries about their child, or videos sent through messaging apps like WhatsApp. When we share things about our children online without involving them in that decision-making process, we're missing out on a valuable opportunity to teach our children and model for our children the idea of consent.



Nuno tried to highlight an important issue, *“There is sometimes a tendency to deprioritise these crimes (AI CSAM) due to resources and lack of clear laws, however even if a real kid hasn't been harmed, we can't just ignore AI-generated child abuse altogether.”*



**What did the audience ask the speakers?**



**How can AI be employed to create digital literacy across populations in the context of how we can avoid desensitising violence?**

• **Answer**

Emma emphasised the need for responsible AI to address the troubling issue of harmful AI. She cited the UK's Stop It Now campaign, which uses AI keyword detection on Pornhub to trigger warnings when users search for illegal content related to child exploitation. This system notifies users and directs them to a helpline, showcasing how AI can prevent harm and offer immediate support.

*Learn more about this campaign: [IWF, Stop It Now, and Pornhub launch first of its kind chatbot](#)*

 Link



• Answer

**How can global best practices be integrated into other countries?**

Shailey stressed that addressing AI-generated child sexual abuse and related harms requires a globally inclusive discussion, bringing together stakeholders from diverse countries to share best practices. She cautioned that solutions effective in one context may not easily translate to others, emphasising the need for context-aware approaches that represent the voices of the global majority, not just those from resource-rich, AI-invested nations. The focus, she argued, must extend beyond regulating AI development to encompass its use and potential for misuse, illustrating the global interconnectedness of the problem with the example of a perpetrator in one country potentially exploiting images from another using AI tools developed elsewhere. Ultimately, the importance of overcoming narrow, nationalistic perspectives to address this complex, global challenge effectively.



**This is a trend that we are seeing here in India, and you know where there is something like a digital arrest that is happening, you know, on a call. There are scammers - not just looking at children, but also adults - where they are scamming you by saying that you know, we found a parcel which has some drugs, or there is something that is there, and they put you on a call throughout, and they make sure that for the next 4 or 5 hours your phone and everything is accessible to them, and they're either. It's a financial scam at this point in time, but definitely, these are also linked to child sexual abuse and exploitation. So in most of these cases, internationally, or probably, whether because you have, you know, exposure to the crimes that are happening across the world - What is the 1st step that a victim can take, or what is the right of a victim when they are facing crimes in the online space, whether it is AI driven or digital crimes?**

## • Answer

In response to the question regarding the best first steps and rights for victims of online crimes, including those driven by AI, Nunu emphasised that the most important action is to report the crime to the police. He stressed that underreporting is an obstacle to addressing online crime effectively, as unreported incidents are often overlooked by law enforcement and policymakers. Nunu acknowledged that victims frequently hesitate to report, believing that police action will be ineffective or wishing to simply forget the traumatic experience. While recognising these understandable victim perspectives, he countered that from a law enforcement and policy standpoint, this underreporting hinders efforts to understand and combat these crimes. For victims dealing with the creation or circulation of harmful images, particularly relevant in the context of AI-CSAM, Nunu pointed to the existence of an image upload portal, though he did not specify the provider. He explained that this portal allows victims to upload harmful images, which are then "hashed" and reported to the Internet and Electronic Service Providers for detection and takedown by online platforms. Nunu underscored key messages for victims: reassurance that 'it's not their fault,' affirmation that 'police are there to help,' and awareness of ongoing efforts to remove harmful images. However, Nunu also asserted that ideally, the onus should not solely be on victims. He proposed that online platforms bear a greater responsibility and should proactively leverage AI technology to detect and remove illegal AI-generated content, moving beyond hash-based detection to identify and prevent the spread of novel harmful material.

## Key Takeaways

This webinar highlighted the rapidly escalating threat of AI-enabled child exploitation and underscored the urgent need for a multi-faceted, collaborative response. Key takeaways from the discussions include:

- **Countering Desensitisation to AI-Generated Harms:** Stakeholders must actively work to combat the desensitisation that can arise from the constant and overwhelming flow of violent and harmful content produced by AI. The sheer scale and realistic nature of AI-CSAM risks normalising this horrific material; it is crucial to prevent this content from being perceived as the new normal and to reinforce the fundamental unacceptability of CSEA in all its forms, including AI-generated versions.
- **Collaboration is Important for Effective Action and Victim Support:** Collaboration across all sectors is not just beneficial but essential to effectively tackle the complex harms of AI-driven child exploitation. This includes law enforcement agencies sharing intelligence internationally, tech companies developing and implementing safety measures and detection tools, educators promoting digital literacy, parents fostering online safety at home, and NGOs providing crucial victim support services. A united front is necessary to combat these borderless crimes and to ensure comprehensive support for victims affected by AI-CSAM and related online harms.
- **Legislation Must Keep Pace with Rapidly Evolving AI Threats:** Legal frameworks are struggling to keep up with the speed of AI development and its misuse for child exploitation. The webinar highlighted the urgent need for legislation to catch up and effectively address AI-driven harms. The EU AI Act was presented as a pioneering example of proactive legislation aimed at governing AI and mitigating its potential harms, offering a potential model for other regions to consider in establishing legal frameworks to combat AI-CSAM.
- **India's I4C: A Model for Collaborative Cybercrime Response:** The I4C exemplifies a proactive and collaborative approach to combating cybercrime, including child sexual abuse material. The webinar showcased I4C's effective strategies in coordinating responses across different agencies and stakeholders, with a strong emphasis on the immediate takedown of CSAM to prevent re-victimisation and limit further harm to children. This model highlights the importance of centralised coordination and rapid response mechanisms in addressing the time-sensitive nature of online child exploitation.
- **Investigations of AI-Driven Harms Must Be Prioritised, Not Deprioritised:** Despite the complexities of investigating and prosecuting AI-CSAM cases and the lack of specific legislation in some areas, law enforcement must resist the temptation to deprioritize these investigations.

- As emphasised, AI-generated harms are real and can have devastating impacts. Deprioritising these cases would send a harmful message and fail to address a growing area of child exploitation in the digital age. Resources and training must be allocated to equip law enforcement to effectively investigate and respond to AI-CSAM, alongside traditional forms of child abuse.
- **AI Literacy and Education are Crucial Across All Stakeholder Groups:** Education and literacy about AI, its potential harms, and online safety are vital for all stakeholders. This includes not just educating children and the general public about online risks like sharenting and AI-CSAM but also equipping law enforcement, educators, tech developers, and policymakers with a deeper understanding of AI technologies, their potential for misuse, and effective mitigation strategies. Widespread AI literacy is essential to building a society that is informed, resilient, and proactive in addressing AI-related harms to children.

Continued dialogue and exchange of learnings and ideas across international borders and stakeholder groups is critical to deepen our collective understanding of both the ways AI can generate harm and the crucial role AI can play in mitigating those very harms, ensuring a safer digital future for children globally.