

CSR



TASI 2025

**Trust And Safety India
Festival**

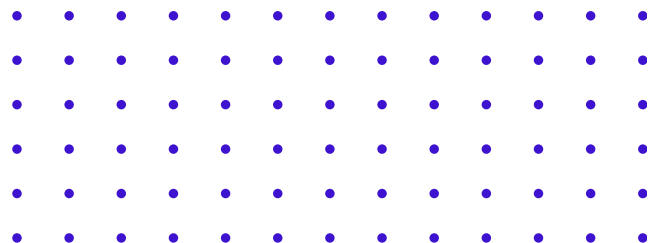


TABLE OF CONTENTS

EXECUTIVE SUMMARY4

EVENT HIGHLIGHTS5

CONTEXT AND OVERVIEW6

INAUGURAL KEYNOTE: DR. S. JAISHANKAR16

EXPERT PANELS19

IMMERSIVE SESSIONS39

WORKSHOPS47

SPOTLIGHT TALKS56

DIPLOMATIC PARTNERS & INTERNATIONAL ENGAGEMENT59

RECOMMENDATIONS66

AS WE LOOK FORWARD: TASI 202669

NOTE OF GRATITUDE72

SPECIAL THANKS73

SUGGESTED READING & MATERIALS75

OUR PARTNERS79

EXECUTIVE SUMMARY

The Trust and Safety India Festival (TASI) 2025, held on October 7–8, 2025, in New Delhi, marked a significant milestone as India's first national convening focused on trust and safety. Distinct in being led by civil society, TASI created a rare, collaborative space for dialogue between government, industry, academia, and civil society. The festival sought to reimagine what it means to build safer, more inclusive, and accountable digital environments in an era shaped by artificial intelligence and rapid technological change.

Recognized as a pre-summit event for India's upcoming AI Impact Summit (February 2026), TASI underscored India's growing role in shaping global conversations on technology governance, ethics, and digital safety. Convened by the **Centre for Social Research (CSR)** and **The Trust & Safety Forum**, with the support of the Ministry of Electronics and Information Technology (MeitY), and in collaboration with leading diplomatic missions, technology platforms, and international organizations, the festival embodied a collective commitment to ensure that digital transformation advances both innovation and the safety and well-being of all users, particularly women, children, and marginalized communities.

The festival comprised 30+ sessions across multiple formats over two days including, panel discussions, workshops, fireside chats, roundtables and interactive spotlights

Programming addressed critical areas at the intersection of technology and society: examining how AI systems can perpetuate or mitigate harm; analyzing platform transparency mechanisms and their effectiveness; exploring vulnerabilities facing children and youth in digital creator economies; documenting the occupational challenges faced by trust and safety professionals; investigating the gendered dimensions of online violence; and conceptualizing digital spaces designed for user care and wellbeing.

The festival generated nuanced takeaways, practice-oriented insights, and research-aligned recommendations, along with commitments for continued engagement beyond the two-day convening. TASI 2025 highlighted India's capacity to lead on trust and safety through approaches rooted in local contexts while remaining in dialogue with global frameworks and partners. The conversations sparked, relationships built, and collaborative commitments made lay the groundwork for ongoing efforts to ensure that India's digital future advances innovation while safeguarding its most vulnerable users.

EVENT HIGHLIGHTS

Date and Venue:

October 7–8, 2025 | The Ambassador, New Delhi

Participation:

Over 300 participants representing 15 countries came together for the inaugural Trust & Safety India Festival (TASI) 2025, a first-of-its-kind platform dedicated to advancing inclusive, ethical, and human-centered digital governance.

Geographic and Sectoral Representation:

TASI 2025 convened a diverse mix of voices, from government and industry leaders to academia, civil society organizations, think tanks, diplomats, and digital rights experts. The event saw active participation from representatives of India's Ministry of Electronics and Information Technology (MeitY) and the Ministry of External Affairs (MEA) alongside international agencies, global platforms, and development organizations.

Sessions:

Across two days, more than 30 panels, workshops, and spotlight sessions unpacked the intersections of AI governance, digital safety, gender, youth protection, and well-being in trust and safety work.

Key Tracks:

AI Governance and Safety:

Building ethical, accountable, and inclusive AI frameworks rooted in transparency and global cooperation.

Child Protection:

Designing safer digital environments for children, grounded in privacy and protection from online harms.

Youth Wellbeing:

Supporting young people's healthy engagement with digital spaces, balancing participation, autonomy, and safety as they navigate online environments.

Gendered and Sexualized Harms Online:

Addressing technology-facilitated gender-based violence (TFGBV), image-based abuse, and emerging forms of online exploitation.

Trust and Safety Workforce Well-being:

Recognizing emotional labor in moderation and safety teams, and prioritizing mental health in digital governance ecosystems.

Safety by Design:

Embedding safety, transparency, and accountability into technologies and AI systems from the outset, ensuring user protection is integral throughout the design and deployment process.

Platform Responsibility and Cross-Sector Collaboration:

Strengthening cooperation among governments, civil society, and platforms to address online harms systematically.

TASI 2025

October 7-8
New Delhi, India

350+

participants

3 days

30+ sessions

100+

speakers

15

countries

Format:

PANELS

WORKSHOPS

PRESENTATIONS

ROUNDTABLES

IMMERSIVE SESSIONS

Sectoral Representation

Topics

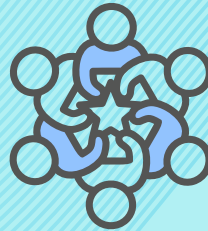
- AI Governance
- AI Safety
- Women's Safety
- Child Safety
- Safety by Design
- Youth Voice
- Mental Health & Wellbeing
- Scams & Frauds
- Women's Online Participation
- Child Influencers & Creators
- Addressing TFGBV
- Digital Rights
- Cross-Sector Collaboration



Government



Platforms



Civil Society



Academia



Diplomatic Missions



Multilateral Orgs

OFFICIAL PRE-SUMMIT EVENT FOR INDIA'S AI IMPACT SUMMIT

TASI 2025 was officially designated as a pre-summit event for India's AI Impact Summit (February 19–20, 2026), establishing trust and safety as foundational pillars of India's AI governance vision. The festival opened with a keynote address by Dr. S. Jaishankar, Minister of External Affairs, whose participation underscored the government's commitment to centering safety, ethics, and inclusion in India's approach to AI.

As one of the key pre-summit convenings contributing to the India AI Impact Summit 2026's focus on People, Planet, and Progress, TASI 2025 created space for critical conversations on digital safety, platform responsibility, child and youth protection, and gendered harms informing the national AI framework ahead of the global summit.

High-level participation from the Ministry of Electronics and Information Technology (MeitY) and NITI Aayog strengthened policy discussions throughout the festival. International missions from France, the Netherlands, Sweden, Canada, and Australia brought valuable global perspectives, while technology companies engaged meaningfully with civil society organizations, academic institutions, and policy experts.

This diverse, multi-stakeholder gathering demonstrated the essential role of inclusive consultation in shaping India's AI governance roadmap—one that aligns government priorities, industry innovation, civil society expertise, and the lived realities of communities most impacted by emerging technologies. As the country moves toward the AI Impact Summit in 2026, TASI 2025 set the tone for a collaborative, safety-centered, and future-ready AI ecosystem.



सर्वजन हिताय | सर्वजन सुखाय
WELFARE FOR ALL | HAPPINESS OF ALL

ABOUT THE CENTRE FOR SOCIAL RESEARCH

The Centre for Social Research (CSR) is a pioneering organization dedicated to advancing gender equality and women's empowerment. CSR has been at the forefront of the women's movement in India, for over four decades and has been working to address various social, cultural, and economic challenges faced by women. The organization engages in research, advocacy, capacity-building, and community outreach initiatives to promote gender justice and create a more inclusive society.

Its areas of focus include combating violence against women; promoting women's political participation and economic empowerment; gender, water and climate change; and advocating for gender-sensitive policies and legislation.

Recognizing the evolving challenges of the digital age, CSR has developed a comprehensive Digital Safety and Online Well-being program that addresses online gender-based violence, cyber harassment, digital privacy, and responsible technology use. Through digital literacy workshops, training sessions, and strategic advocacy efforts, the organization works to empower women and marginalized communities to navigate digital spaces safely. CSR partners with educational institutions, corporates, and civil society organizations to conduct awareness campaigns, develop policy recommendations, and build capacity among diverse stakeholders, from young women and adolescents to educators and law enforcement. This work has positioned CSR as a leading voice in advocating for digital rights and gender-responsive technology policies in India.

CSR, through its collective efforts, aims to create a future where all genders in India enjoys equal rights, opportunities, and the agency to drive positive change.



ABOUT THE TRUST AND SAFETY FORUM

The Trust & Safety Forum is a global platform dedicated to advancing digital safety through collaboration among technology companies, policymakers, civil society organizations, and researchers. Co-founded by Caroline Humer and Jean-Christophe Le Toquin, the festival brings together diverse stakeholders to address critical challenges in creating safer online environments.

TASI 2025 was co-organized by the Centre for Social Research (CSR) and the Trust & Safety Festival, marking the festival's first edition in India. This collaboration brought together CSR's four decades of leadership in gender equality and digital safety with the Trust & Safety Festival's global network, creating a landmark event in New Delhi that centered Indian perspectives and priorities in shaping the future of online safety for the Global South.



“

This is the first time global and local platforms, tech providers, civil society and public officials gather in a leading Global South country to advance Trust & Safety. A different perspective for a borderless dialogue, and much awaited.

Jean-Christophe Le Toquin
Co-Founder, Trust & Safety Forum



“

The Trust & Safety Festival India is more than just an event—it marks a pivotal moment in the global online safety landscape. For the first time, voices from the Global South are shaping the future of digital trust, addressing challenges like AI-driven scams, gender-based violence, child protection, and responsible innovation with Indian insights and global relevance.

Caroline Humer
Co-Founder, Trust & Safety Forum

TEAM TASI

TASI 2025 came together through the dedicated efforts of an organizing team committed to advancing trust and safety dialogue in the Global South.



Dr. Ranjana Kumari
Director
Centre For Soical
Research



Jyoti Vadehra
Head, Online Safety
and Wellbeing, Centre
for Social Research



Karuna Nain
Online Safety Expert/
Advisor, Centre for
Social Research



Saquib Jamil
Project Coordinator -
Media and
Communications,
Centre for Social
Research



Akankshya Acharya
Project Coordinator -
Digital Safety and Online
Wellbeing, Centre for
Social Research



Ketaki Ruchika Singh
Project Coordinator -
Media and
Communications,
Centre for Social
Research

FOREWORD

The Centre for Social Research's engagement with the trust and safety ecosystem is not a divergence from our longstanding mission; it represents its logical and necessary extension. For more than four decades, we have worked to advance gender equality, strengthen child protection systems, and promote inclusive social development. As digital transformation accelerated across India and the Global South, it became increasingly clear that the patterns of harm emerging online are closely connected to the structural inequalities we have addressed offline. The digital sphere, though borderless in reach, mirrors and at times intensifies the existing vulnerabilities of women, children, and marginalized communities.

The global digital environment today is defined by profound dualities. Technology has expanded learning opportunities, enabled financial inclusion, and offered new pathways for expression and participation. At the same time, it has amplified longstanding forms of discrimination, exclusion, and gender-based violence in new formats and at greater speed. Digital platforms often reflect social realities rather than create them: the same barriers that limit women's mobility offline can reappear through online harassment; risks faced by children in physical spaces can surface through exposure to harmful content or unsafe interactions. Understanding this continuity allows us to approach trust and safety not as a reaction to emerging threats, but as an essential extension of broader social equity and rights-based efforts.

Behind each policy debate lie human experiences.

Young women who have faced harm when private information was misused; families who have struggled after falling prey to online financial fraud; and children navigating emotional distress resulting from sustained online bullying. These are not isolated incidents but part of a wider pattern that underscores the need for responsive, accountable, and survivor-centered digital ecosystems.

It is in this context that CSR's digital safety work emerged. Yet we have always recognized that safeguarding digital spaces cannot be achieved by any single institution. Technology companies possess technical capacity but cannot capture every cultural nuance. Governments provide regulatory frameworks, yet depend on community-level insights to ensure these frameworks are equitable and effective. Civil society organizations bring lived experience, research, and advocacy, but require constructive collaboration with all stakeholders to drive systemic change. Trust and safety therefore demands a multistakeholder model grounded in shared responsibility and collective action.

At the global level, we observed a critical gap: while international dialogues on trust and safety are expanding, the voices, priorities, and contextual realities of the Global South remain insufficiently represented. Countries like India, which host vast digital populations with diverse socio-economic contexts, face unique challenges and offer innovations that must inform the global policy landscape. This recognition guided the creation of the Trust and Safety India Festival – TASI 2025.

TASI 2025 became India's first civil society-led platform dedicated exclusively to trust and safety. It convened government representatives, international experts, technology platforms, academics, and civil society leaders to build shared understanding and identify actionable pathways. With over 500 participants from 15 countries, the convening demonstrated the global resonance of India's experience and the urgency of placing safety, rights, and inclusion at the center of digital governance.

What set TASI apart was its integrated perspective. Discussions bridged AI governance, platform accountability, child online protection, digital gender-based violence, and the ethics of emerging technologies. The deliberations emphasized a people-centered vision, one aligned with international human rights standards and inclusive of cultural, socio-economic, and gendered realities.

As we advance toward TASI 2026 and strengthen our engagement with global partners, CSR remains committed to ensuring that digital governance frameworks reflect the lived experiences of communities, particularly those historically marginalized. We will continue to foster collaboration between governments, technology companies, multilateral agencies, and civil society to shape a digital order rooted in equity, accountability, and safety. We will further amplify the perspectives of India and the Global South in international policy discourse, contributing to a balanced and context-sensitive global architecture for trust and safety.

TASI is not merely an annual event; it is a sustained commitment to building resilient digital ecosystems that protect individuals, support those at risk, and ensure that technological progress serves the public good. As we move into the next phase of this work, our resolve remains unwavering. Together with our partners worldwide, we aim to shape a digital future where safety is guaranteed, trust is strengthened, and every person can participate fully and confidently.

Dr. Ranjana Kumari

Director, Centre for Social Research

1. INTRODUCTION

1.1 About TASI 2025: Why This Festival

In an age where nearly every aspect of life intersects with the digital world, trust and safety have become essential to how people connect, express themselves, and access opportunity. From protecting children online to building ethical AI systems, from ensuring platform accountability to defending digital rights, the landscape of online safety is vast, urgent, and constantly evolving.

Across the world, these conversations are gaining momentum, yet many remain shaped by frameworks and experiences from the Global North. There is a critical need to elevate voices, solutions, and leadership from the Global South, to ensure that safety frameworks

are not only scalable but also contextually grounded, culturally aware, and representative of the lived realities of diverse digital communities.

The Trust & Safety India Festival (TASI) 2025 emerges from this imperative. Conceived as the first civil society-led convening of its kind in India, the Festival brings together an unprecedented mix of actors from governments and tech platforms to civil society, academia, and multilateral organisations. It builds on the energy of the Trust & Safety Forum in Europe, but firmly roots the conversation in the Indian and Global South context, where digital inclusion, innovation, and inequality intersect in unique ways.



1.1.1 The Vision: “Building Safer Digital Ecosystems Through Shared Responsibility”

TASI 2025 is guided by a shared vision: “building safer digital ecosystems through shared responsibility.” It recognises that no single actor, whether platform, policymaker, or community, can achieve safety alone. True progress lies in collective ownership, in aligning our approaches and learning from one another’s experiences to create an internet that enables trust, empathy, and empowerment for all.

1.1.2 The Format: Multi-Stakeholder, Cross-Sector, Collaborative

Over two days, TASI 2025 blends the rigour of policy dialogue with the energy of collaboration and creativity. Through high-level plenaries, interactive workshops, innovation showcases, and networking receptions, participants explore issues such as AI ethics, online violence, digital citizenship, and cross-border governance. Each session is designed not merely for discussion, but for action and partnership-building, laying the groundwork for concrete frameworks, pilot collaborations, and long-term initiatives.

“To address Technology Facilitated Gender Based Violence, there is a need for a multipronged approach by improving digital literacy and security; improving content moderation and response in collaboration with technology companies; strengthening laws and building capacities of law enforcement officials; investing in survivor-centric resources and support; and addressing social norms that perpetuate violence.”

Ms Kanta Singh
Country Representative, UN Women India

1.1.3 The Spirit: Learning, Listening, and Co-creation

At its heart, TASI is more than a conference, it is a festival of ideas, listening, and co-creation. It invites participants to bring their expertise and lived experience into a shared space of mutual respect and discovery. It is where regulators learn from youth advocates, engineers from survivors, and researchers from community organisers. This spirit of learning and inclusion is what makes TASI both unique and necessary, a place where trust is not just discussed, but actively practiced.

1.2 Why Now

India today stands at the centre of the global digital conversation, home to one of the largest online populations and a rapidly expanding tech ecosystem. Yet it also mirrors the world’s most pressing challenges: misinformation, online abuse, algorithmic bias, and disparities in access and safety. The time has come for India to lead, not only as a market, but as a thought partner and innovator in reimagining what a safer digital future can look like. TASI 2025 is that platform. It is where ideas become shared commitments, and where building safer digital spaces becomes a collective mission, one grounded in empathy, equity, and collaboration.

“As one of the largest and fastest-growing digital populations, India is a natural centre for this gathering of key stakeholders and leading trust and safety experts. These crucial conversations will shape the future of the internet, making it a safer place for all and it is a privilege to be a part of it.”

Sophie Mortimer
Manager, UK Revenge Porn Helpline, SWGfL

2. KEYNOTE ADDRESS

Dr. S. Jaishankar, Minister of External Affairs

Dr. S. Jaishankar, Minister of External Affairs, delivered the inaugural address, articulating how technology is reshaping global relations and domestic governance, and how India envisions a digital future rooted in inclusion, accountability, and trust. His remarks positioned trust and safety not as peripheral concerns, but as core pillars of India's engagement with the digital world.

Opening Keynote Transcription

Dr. Ranjana Kumari ji,

Friends,

It is a great pleasure to join you all this morning as we commence the Trust and Safety India (TASI) Festival. As a pre-Summit event for the February 2026 AI Impact Summit, it kicks off a set of conversations that are of great interest both in India and to the world. I commend the Centre for Social Research India and you, Dr. Ranjana Kumari ji personally, for taking this initiative.



2. Friends, in human history, the advancement of progress has been determined by the march of technology. Yet, that has not always been a linear path. The promise and the perils have always been two sides of the same coin. Choices made in the deployment and application decided the difference between empowerment and exploitation, between democratization and domination and between partnership and polarization. Today, we are at the cusp of an immense change and decisions made by us, individually and collectively, will help decide the fate of the foreseeable future.

3. Now, if any of you think that this is an exaggeration, consider what is likely to happen in just the next few years. AI will transform our economies, it will change our work habits, it will create new health solutions, it will enhance educational access, it will improve efficiencies and it could even lead to a new lifestyle. The difference with the past is that this particular transformation that is going to come upon us, this will be all-pervasive and not just sectoral. It will impact every citizen in every corner of the world. New efficiencies and new possibilities will certainly emerge; but so too will new players and new power centres. For this very reason, we need to adopt a balanced approach to the governance of AI and to ensure that there are adequate guardrails in place to protect digital nagriks. As Prime Minister Narendra Modi has noted: "Technology is a force for good, but only if humanity guides it".

4. Friends, Trust and Safety are essential for the implementation of Responsible AI. For a society like India, this means development of indigenous tools and frameworks, self-assessment protocols for innovators and establishing relevant guidelines. Only then can

we be assured that the development, deployment, use and governance of AI is safe and accessible. India bears a special responsibility because many other nations – especially those of the Global South – they look to us for inspiration, and certainly in the case of the digital public infrastructure, today what we have achieved in the last decade, if you see today the scale of delivery, the improvement in governance, the efficiencies with which we serve the public, I think it is already resonating across the world. As a Foreign Minister, when I go abroad, this is often the topic of conversation, and I can see very much, this being extrapolated into the world of AI as well.

5. Now, different societies have put different levels of emphasis on the benefits and on the risks of AI. Understandably, some of the narrative is influenced by those who have skin in the game. But at the end of the day, it is imperative that we take a sober and we take a balanced view. After all, concerns about bias, ethics, privacy and vulnerability are well-founded, deriving from pre-existing experiences. We also need to take particular precaution against the danger of losing trust in institutions and practices that underpin our daily life. Already, we see an inclination to question the umpire or even the playing field when results don't go in our favour. Maintaining – and I would argue in the era of AI – even strengthening trust therefore is of utmost importance.

6. Friends, encouraging a global discourse on AI is obviously essential. Indeed, every time a game changing technology has appeared on the horizon, it has triggered off similar endeavours. But history is also a witness to the reality that those exercises have not been easy.



There is the obvious temptation to put narrow self-interest above collective good. And that is why we need to rise to this particular challenge, because what is at stake cannot be over-stated. It is not just the ambitions of polities but the personal interest of every citizen on this planet.

7. Now we, in India, we have consistently advocated for global AI governance and the shaping of an international agenda. We have demonstrated this during our G20 Presidency and strongly urged its usage for advancing Sustainable Development Goals, while safeguarding trust, safety, fairness and accountability. As a founding member of the Global Partnership on AI, we promoted the New Delhi Declarations which envisaged responsible and inclusive AI. We participated in the AI Summits in Bletchley Park and Seoul

and co-chaired the Paris AI-Action Summit last year. The AI-Impact Summit in 2026, with a focus obviously on impact, will take this forward.

8. So let me conclude by saying this - this is an endeavour whose success depends on our ability to ensure multi-stakeholdership. And that is why the TASI Festival today is important because it sends out a strong inclusive and pro-people message on this critical issue. I thank the organizers once again for inviting me and wish your deliberations all success.

Thank you very much.

New Delhi
October 07, 2025

3. EXPERT PANELS

3.1. Panel: From Paris to Delhi: Shaping Global AI Governance and Ethical Innovation

Date: October 7, 2025

Time: 09:15 AM

Venue: Dome, Taj Ambassador, New Delhi

Panelists:

- Sunil Abraham – Public Policy Director, Data Economy and Emerging Tech, Meta India
- Abhishek Singh – Additional Secretary, Ministry of Electronics and Information Technology, Government of India
- Nandita Baruah – Country Representative, India, The Asia Foundation
- Rajesh Ranjan – India Lead, Core Government Affairs and Public Policy, Google

- Moderator: Rohit Kumar – Founding Partner, The Quantum Hub

Key Discussions:

The first panel of TASI 2025, “From Paris to Delhi: Shaping Global AI Governance and Ethical Innovation,” set the tone for the two-day festival with a high-level dialogue on India’s growing leadership in responsible and ethical AI. Abhishek Singh outlined India’s vision for inclusive, human-centric AI grounded in trust, fairness, and safety. He announced the establishment of the AI Safety Institute, envisioned as a hub for developing tools on bias mitigation, watermarking, deepfake detection, and ethical certification, positioning India as a key global collaborator on AI safety and



governance. Sunil Abraham discussed Meta's "human-in-the-loop" model of innovation, emphasizing the balance between open science and context-specific safeguards, while Rajesh Ranjan highlighted Google's initiatives including Project Vani, Project Amplify, and Synth ID, focused on transparency, local language inclusion, and digital accessibility. Nandita Baruah underscored the vital role of civil society in bridging policymakers and communities to ensure that AI governance frameworks reflect lived realities and prevent the reinforcement of social inequities.

The panel collectively emphasized that AI governance cannot operate in silos – it must be participatory, globally aligned, and rooted in societal ethics. Rohit Kumar concluded by observing how global discussions on AI have evolved from questions of safety and frontier models to inclusion and impact, noting that India's approach should not only align with global frameworks but also lead by embedding democratic values, diversity, and accountability at its core.

Takeaways:

- India's AI Safety Institute will serve as a hub for research, red-teaming, and cross-sector collaboration,
- Promoting ethical, inclusive, and resilient AI ecosystems.
- Civil society participation is essential for bridging technological design and community realities, ensuring that AI policies reflect human rights and social equity.
- Context-aware governance, blending industry innovation, localized regulation, and open-source collaboration, is vital for shaping safe, globally relevant AI models.

Recommendations:

- Advance a collaborative AI Safety Institute that brings stakeholders together to develop shared tools for bias mitigation, ethical assurance, and watermarking.
- Support the integration of ethical, transparent, and inclusive design practices that reflect India's social diversity.
- Build India-led Global South collaborations that promote equitable access, shared governance, and culturally grounded AI innovation.



3.2. Panel: Panel: AI for Safety

Date: October 7, 2025

Time: 11:00 AM

Venue: Dome, Taj Ambassador, New Delhi

Panelists:

- Jordan Benavidez - Director, Safety by Design, Match Group
- Henry Adams - Director, Trust & Safety, Resolver
- Akash Pugalia - Chief Digital Officer, TP
- Snigdha Bharadwaj - Director, Google Generative AI, Trust & Safety
- Moderator: JC Le Toquin - Co-Founder, Trust & Safety Festival

Key Discussions:

The panel explored AI's transformative role in digital safety ecosystems, balancing its potential risks and solutions. Jordan Benavidez detailed Match Group's "Safety by Design" approach, focusing on inclusive product frameworks that prioritize eliminating child harm, harassment, and hoaxes. Akash Pugalia emphasized AI-powered "human-in-the-loop" detection models that combine machine speed with the nuance of human ethics to better spot emerging harms in real time. Henry Adams provided historical perspective, reminding attendees that generative AI is rooted in decades of research, but human oversight remains crucial to prevent bias and uphold cultural sensitivities. Snigdha Bharadwaj highlighted India's collaborative efforts using multilingual AI and watermarking technologies to ensure digital

authenticity and trust, including projects like Project Vani and the Bhashini platform. JC Le Toquin stressed that such innovations must be embedded within multi-stakeholder governance frameworks to ensure trust and accountability at scale.

Takeaways:

- AI must augment human judgment rather than replace it to maintain ethical oversight.
- Transparency and content provenance tools like watermarking are essential, especially in diverse linguistic environments.
- Multi-sector collaboration fosters robust AI safety practices customized to regional cultures.
- Create cross-sector governance frameworks incorporating local context and global best practices.



3.3. Panel: Protecting Children Online

Date: October 7, 2025

Time: 12:00 PM

Venue: Dome, Taj Ambassador, Delhi

Panelists:

- Shilpi Singh - Member, Gurgaon Moms
- Co-Founder, Studio4, Unhotel Global
- Nikhil Taneja - Co-Founder & Chief, Yuva
- Uthara Ganesh - Head of Public Policy, India & South Asia, Snap Inc.
- Natasha Jog - Head, Public Policy, Meta India
- Moderator: Kazim Rizvi - Founding Director, The Dialogue

Key Discussions:

The session underlined the multifaceted challenge of online child safety where parental roles, platform accountability, and regulatory oversight intersect. Shilpi Singh shared emotional accounts about AI-generated harmful content and the ever-growing complexity of digital parenting.

Nikhil Taneja explained how algorithms may inadvertently funnel children to unsafe spaces, advocating AI as a “second verifier” for content safety beyond user reports. Uthara Ganesh highlighted Snap’s deployment of AI-based parental controls and wellbeing nudges, emphasizing user education alongside technology solutions. Natasha Jog focused on embedding privacy-by-default and age-gating as crucial platform features to mitigate risk. Kazim Rizvi emphasized that effective solutions require collective action from families, regulators, and platforms.

Takeaways:

- Platforms must embed privacy and safety defaults tailored for minors.
- Parental digital literacy and emotional support play a critical role alongside technological safeguards.
- Promote greater transparency from platforms through voluntary reporting and shared best practices on child safety metrics.



Recommendations:

- Encourage child online safety standards that prioritize privacy-first design and age-appropriate experiences.
- Promote voluntary transparency efforts where platforms share insights and best practices on child safety metrics.
- Strengthen parental education initiatives that support digital wellbeing and emotional safety.

3.4. Panel: CTRL + SHIFT + RESPECT: Building Safer Digital Spaces for Girls

Date: October 7, 2025

Time: 2:15 PM

Venue: Dome, Taj Ambassador, New Delhi

Panelists:

- Akansha Kasera - Senior Market Engagement Manager, GSMA
- Shireen Vakil - Child Rights Advocate
- Sadhna Singh - Consultant, NITI Aayog

- Josephine van Zanten - Second Secretary Political Affairs & Human Rights, Netherlands Embassy
- Japleen Pasricha - Founder-CEO & Editor-in-Chief, Feminism in India
- Moderator: Kavita Ayyagiri - Country Director, Girl Effect India

Key Discussions:

The discussion illuminated the pervasive gendered digital harms which mirror broader social inequities. Panelists discussed structural barriers including victim blaming, policing of women's voices, and enforcement gaps in digital abuse cases. Akansha Kasera described how patriarchal norms silence women online, linking cultural misogyny with platform moderation challenges. Shireen Vakil cited legal ambiguities and under-reporting as barriers to justice for survivors. Sadhna Singh highlighted digital illiteracy and shared phone use compromising women's privacy in rural areas. Josephine van Zanten emphasized governance blending regulation with



empowerment. Japleen Pasricha raised the pressing need for gender-disaggregated data to effectively tailor policies. Kavita Ayyagiri concluded by stressing preventive design embedding respect into online ecosystems.

Takeaways:

- Online gender violence needs intersectional policy and technological strategies.
- Closing digital literacy and access gaps is critical for marginalized women.
- Proactive, respect-based platform design offers sustainable safety solutions.

Recommendations:

- Roll out a Digital Gender Equity Charter with focus on gender impact assessments.
- Strengthen cyber law enforcement focused on gender-based violence.

- Invest in community-driven digital literacy for women in underserved areas.

3.5. Panel: Click to Protect - How Digital Platforms Can Disrupt Human Trafficking Networks

Date: October 7, 2025

Time: 3:15 PM

Venue: Dome, Taj Ambassador, New Delhi

Panelists:

- Neda Niazian - Group Director, Trust & Safety, Booking.com
- Amelia Wierda - Legal Counsel for Human Rights, Booking.com
- Nick Dale - Director of Intelligence, Stop the Traffik
- Lena Dasgupta Basu - Programme Manager, Child Protection, Child in Need Institute (CINI)
- Moderator: Hasina Kharbhih - Founder, Impulse NGO Network



Key Discussions:

Panelists exposed how human trafficking increasingly exploits digital invisibility, demanding innovative technological and cooperative responses. Neda Niazi explained how Booking.com uses pattern detection and cross-platform intelligence sharing to identify trafficking activities. Nick Dale emphasized the linkage of digital signals with actionable law enforcement intelligence, while maintaining survivor privacy and trust. Lena Dasgupta Basu advocated for youth involvement and community empowerment as key factors in safety-building. The group called for a “Digital Covenant” to bind platforms to humanitarian accountability principles.

Takeaways:

- Interoperable data-sharing and coordinated intelligence are key to disrupting trafficking.

- Survivor anonymity encourages reporting and intervention.
- Community-based youth engagement strengthens protective ecosystems.

Recommendations:

- Explore a shared digital repository to support secure, multi-agency coordination on anti-trafficking efforts.
- Promote cross-industry collaboration to improve trafficking risk detection and safeguard users.
- Expand youth- and community-driven digital safety initiatives that build awareness and resilience.

3.6. Panel: Disconnecting Fraud and Enhancing Safety

Date: October 7, 2025

Time: 4:15 PM

Venue: Dome, Taj Ambassador, New Delhi



Panelists:

- P. K Singh - Additional Secretary, TRAI
- ACP Harmeet Randhawa - Delhi Police, Crime & Technology Focal Special Cell
- Yannick Ragonneau - Cybersecurity Thematic Coordinator of EU project ESIWA
- Vidushi Chaturvedi - Lead Architect, Frontier Tech Hub, NITI Aayog
- Seema Jindal - Head, Public Affairs, Truecaller
- Suhel Daud - Legal Attache, U.S. Embassy New Delhi
- Moderator: Sumeysh Srivastava - Associate Director, The Quantum Hub

Key Discussions:

The panel explored the evolving nature of fraud in India's digital and telecom environments. The TRAI Additional Secretary reviewed enforcement challenges under telecom commercial communication regulations and the need for increased industry compliance. ACP Harmeet Randhawa discussed gaps in police cybercrime capabilities and the urgency of multi-agency collaboration. Yannick Ragonneau shared European best practices on fraud prevention emphasizing traceability and preemptive suspension of suspicious actors. Vidushi Chaturvedi highlighted emergent AI deepfake fraud threats and called for transparent AI tools limiting false positives. Seema Jindal explained Truecaller's crowd-sourced fraud detection innovations. Suhel Daud underlined the necessity of persistent vigilance and real-time intelligence exchange beyond technological measures.

Takeaways:

- Effective fraud control depends on seamless collaboration across industry and law enforcement.
- Transparent AI fraud detection inspires trust and reduces wrongful blocking.
- Awareness campaigns must integrate behavioral insights with legal education.

Recommendations:

- Build momentum toward a unified fraud intelligence grid that connects telecom, finance, and enforcement partners for faster threat response.
- Advance shared approaches to explainable AI so fraud-detection systems remain transparent, trusted, and user-centric.
- Grow cyber-fraud education efforts that blend tech, legal, and behavioural literacy for stronger public awareness.

3.7. Fireside Chat: Dating & Safety – Navigating Trust in Modern Relationships

Date: October 8, 2025

Time: 9:00 AM

Venue: Dome, Taj Ambassador, New Delhi

Speakers:

- Yoel Roth - Senior Vice President & Head of Trust & Safety, Match Group
- Dr. Ranjana Kumari - Director, Centre for Social Research, India
- Moderator: Nilanjana Bhowmik, Journalist

Key Discussions:

The fireside chat explored how digital dating is reshaping trust, intimacy, and safety in India. Yoel Roth highlighted the dual nature of dating apps expanding connection while exposing users to risks such as impersonation, harassment, and coercive control. He outlined Match Group's safety efforts including identity verification, AI-driven detection of harmful behaviors, and proactive in-app safety cues.

Dr. Ranjana Kumari emphasized that despite available reporting pathways, many survivors, particularly women, avoid formal complaints due to shame, stigma, and fear of institutional failure. She stressed the need for trauma-informed communication and empathetic support systems that prioritize survivor safety over procedural compliance.

The role of AI was discussed as both a preventive tool and a limitation: algorithms can detect red flags but cannot replace human

judgment or empathy. The speakers also pointed to gaps in India's social context, limited sex education, normalization of victim-blaming, and low awareness of consent arguing that safety features must be complemented by broader cultural and educational interventions.

Roth and Bhowmik noted that dating platforms increasingly function as social spaces, and therefore have a responsibility to embed safety into every aspect of user experience online and offline.

Takeaways:

- Safety must be built into dating apps, with AI supported by empathetic human review.
- Trauma-informed, stigma-aware systems are crucial to encourage reporting.
- Education on consent and respectful communication is key to digital relationship safety.



Recommendations:

- Co-create Safety-by-Design principles that address risks unique to dating and relationship platforms.
- Require survivor-centric reporting with trauma-informed protocols and clear follow-up.
- Integrate digital consent and relationship literacy into school and university curricula to strengthen preventive safety.

3.8. Panel: Designing for Youth - Privacy, Protection, and Participation in the Digital Age

Date: October 8, 2025

Time: 10:30 AM

Venue: Dome, Taj Ambassador, New Delhi

Panelists:

- Mitthat Hora - Program Officer, British Asia Trust
- Siddharth P. - Co-Founder & Director, RATI Foundation

- Jeff Wu - Co-Founder & Chief Safety Officer, K-ID
- Vaishnavi J - Founder, VYS
- Moderator - Jyoti Vadehra, Head of Digital Trust & Safety, CSR India

Key Discussions:

The panel examined how design choices shape young users' experiences in digital spaces. Siddharth P. presented research on unique Indian digital behaviors, including device-sharing among family members, household permission-seeking patterns, and widening class divides in digital access. Mithhat Hora shared findings revealing that while 99% of children have device access, most lack clear pathways to report abuse, with over 30% admitting to talking with strangers online and 22% meeting them in person. Jeff Wu critiqued age-gating mechanisms, noting that children learn to misstate their age before



understanding why limits exist, and argued for age-appropriate design frameworks that differentiate experiences rather than imposing blanket restrictions. Vaishnavi J. called for Safety by Design as the cornerstone of app development, emphasizing that smaller Indian companies often lack dedicated trust and safety infrastructure. Jyoti Vadehra framed the discussion around design having moral consequences, stressing that youth safety must be an active design principle rather than a reactive patch.

Takeaways:

- Digital policies must account for Indian-specific behaviors like device-sharing and household negotiation dynamics rather than assuming Western single-user models.
- Corporate redressal mechanisms need significant improvement, with platforms taking over ten days to act on serious complaints.
- Age-appropriate design should create differentiated experiences rather than

- excluding youth from safe online environments.
- Safety by Design requires in-app SOS buttons, contextual help mechanisms, and reporting pathways tailored for teen users.

Recommendations:

- Strengthen age-assurance methods that balance privacy with age-appropriate access.
- Improve response pathways for youth safety concerns through clearer, faster escalation.
- Shape adaptable Safety by Design guidance for both large and small companies.
- Develop frameworks that support safe, confident digital participation for young people.

3.9. Panel: Child Influencers in the Creator Economy

Date: October 8, 2025

Time: 12:15 PM

Venue: Dome, Taj Ambassador, New Delhi



Panelists:

- Anshul Tewari - Founder & CEO, Youth Ki Awaaz
- Kriti Trehan - Founder, Data & Co - Law & Policy Advisors
- Manisha Kapoor - CEO & Secretary-General, The Advertising Standards Council of India
- Anupriya Kapur - Digital Creator & Influencer
- Moderator: Barkha Dutt - Journalist & Media Entrepreneur

Key Discussions:

The panel tackled the commercialization of childhood in digital spaces. Barkha Dutt opened by questioning what happens when childhood itself becomes content, arguing against direct monetization of child-created content while supporting creativity without commodification. Anshul Tewari described content creation as children "coding themselves," but warned of psychological risks from digital visibility including comparison and self-harm triggers. Manisha Kapoor highlighted the regulatory void in India, noting that validation through likes and followers has become "the new report card," with concerns around body image distortion, identity formation pressures, and parent-driven commercialization. Kriti T. emphasized that content regulation is inherently subjective and advocated for guardianship models alongside age-appropriate design frameworks. The discussion acknowledged that notions of "safe visibility" differ drastically by class and geography in India, with viral content representing social mobility for some families and social risk for others.

Takeaways:

- There is no comprehensive checklist for parents on what constitutes ethical exposure of children online.
- Accountability must be shared between parents, platforms, regulators, and society rather than resting with one stakeholder.
- Digital empathy requires teaching both children and adults about the permanence of online footprints.
- Banning technology isn't the solution; building ethical frameworks that prioritize children's rights over metrics is essential.

Recommendations:

- Strengthen age-appropriate design practices for child-facing platforms.
- Define clear, practical guidelines for ethical online exposure of children.
- Expand parent education through community and school-based programs.
- Ensure platform features and incentives are aligned with children's rights and healthy digital experiences.



3.10. Fireside Chat: Advancing Women's Voices in Public Life

Date: October 8, 2025

Time: 02:00 PM

Venue: Dome, Taj Ambassador

Speaker: Bansuri Swaraj, Member of Parliament, Government of India

Moderator: Jyoti Vadehra, Head of Digital Trust & Safety, CSR India

Key Discussion:

This fireside conversation examined what it means for women to claim space and leadership in India's public life. Bansuri Swaraj reflected on her journey in law and public service, noting how women often navigate heightened scrutiny and gendered expectations despite rising representation. She emphasized that leadership must be recognized not only through formal titles but also through the lived realities women bring into governance.

A significant portion of the discussion focused on digital participation. Bansuri described the internet as the new public square, one that amplifies women's voices but also exposes them to targeted harassment and misogyny. She called for stronger online safety frameworks, consistent platform enforcement, and redressal systems that protect women who engage in public discourse. She underscored that visibility is essential for democratic participation and urged for systems that support women's safe engagement online.

The conversation also highlighted the need for structural support mechanisms. Bansuri advocated for mentorship networks, community education, and intentional documentation of women's achievements to normalize female leadership. She shared examples of women across professions—law, journalism, teaching, grassroots organizing—whose contributions often remain unrecognized but shape India's democratic fabric.



Takeaways:

- Women's leadership strengthens governance through empathy and lived experience.
- Online participation must be safeguarded from gender-based abuse and intimidation.
- Mentorship, representation, and institutional support are vital for sustaining women in public life.
- Visibility and storytelling help normalize women's leadership across sectors.

Recommendations:

- Create formal mentorship networks for aspiring women leaders across governance and public service.
- Enforce robust online safety norms and transparent redress mechanisms protecting women in digital public spaces.
- Launch national campaigns documenting and amplifying women's contributions to public life to shift societal narratives.

3.11. Panel: Responding to Technology-Facilitated Gender-Based Violence (TFGBV)

Date: October 8, 2025

Time: 2:30 PM

Venue: Dome, Taj Ambassador Hotel

Panelists:

- Barsha Chakraborty - Feminist Practitioner & Digital Safety Trainer
- Shanley Clemot McLaren - Founder, #StopFisha and Global Gender Digital Rights Advocate
- Steven Biddle - Minister Counsellor, Department of Home Affairs, Australian High Commission
- Moderator: Caroline Humer - Co-Founder, Trust & Safety Forum



Key Discussions:

The panel addressed TFGBV as a continuum of offline inequality where patriarchal structures, anonymity, and algorithmic amplification collide. The panel highlighted that over 90% of TFGBV data across South Asia remains unreported or unstructured. Barsha Chakraborty emphasized that reporting tools are designed for systems rather than survivors, calling for co-created redressal processes that avoid retraumatization. Steven Biddle shared Australia's Safety-by-Design framework as a model for preventive architecture and survivor-first solutions. Shanley Clemot McLaren discussed her successful campaign criminalizing non-consensual intimate image sharing in France, calling for feminist data governance and cross-border survivor coalitions. The panel also explored how privacy is understood unevenly in India, emphasizing the need to localize privacy laws so individual autonomy becomes the core of digital safety.

Takeaways:

- TFGBV requires cultural transformation, not just content takedown.
- Safety must be a founding design principle baked into platform architecture, not an afterthought.
- Digital violence doesn't respect borders; global harm requires global response mechanisms.
- Privacy in India is often a privilege of those who can afford it and must be reframed as a right to be guaranteed, not negotiated.

Recommendations:

- Strengthen coordination across law enforcement, women's bodies, and platforms to address TFGBV.
- Embed Safety-by-Design with faster, survivor-centered reporting.
- Advance cross-border cooperation and feminist-aligned data governance.
- Localize privacy efforts to center individual autonomy and informed choice.

3.12. Fireside Chat: Balancing Innovation & Safety – A Dialogue with S. Krishnan

Date: October 8, 2025

Time: 3:15 PM

Venue: Dome, Taj Ambassador

Speakers:

- S. Krishnan - Secretary, Ministry of Electronics & Information Technology (MeitY), Government of India
- Moderator: Karuna Nain, Online Safety Expert and Advisor to Centre for Social Research

Key Discussions:

This fireside conversation offered an inside view of how India is shaping a digital ecosystem where innovation and safety reinforce each other. S. Krishnan outlined India's approach to responsible tech

development, emphasizing that risks today are systemic rather than isolated, intersecting AI governance, user rights, and platform accountability. He highlighted the Digital Personal Data Protection Act and emerging Safety by Design frameworks as foundations for India's next phase of digital governance.

Discussing the balance between innovation and regulation, Krishnan noted that India cannot afford to stifle creativity, but must also prevent digital harms from compounding as technologies scale. He described MeitY's work on anticipatory governance models that allow for experimentation with appropriate guardrails. Krishnan also underscored the importance of supporting startups with simplified safety toolkits and shared compliance resources, recognizing their unique understanding of India's linguistic and cultural diversity.

On youth safety, he flagged the need for age-appropriate protections and greater digital

literacy in education systems, noting gaps in awareness and redressal mechanisms. Addressing AI governance, he stressed the need for transparency, auditability, and clearer definitions of algorithmic harm. Moderating the conversation, Karuna Nain steered the discussion toward global alignment, with Krishnan highlighting India's opportunity to demonstrate a human-centered governance model at the 2026 AI Impact Summit.

Takeaways:

- Innovation and safety must advance together through proactive, collaborative governance.
- Startups require context-sensitive compliance and safety support to scale responsibly.
- Algorithmic transparency and auditability are essential for trust in AI systems.
- Youth safety demands integrated digital literacy and coordinated ecosystem action.



Recommendations:

- Build anticipatory governance approaches that guide responsible AI innovation.
- Improve clarity and evaluability of high-impact AI systems across sectors.
- Co-develop safety toolkits and capacity-building programs for startups through multi-sector partnerships.

3.13. Panel: Well-being for Trust & Safety Workers

Date: October 8, 2025

Time: 3:40 PM

Venue: Dome, Taj Ambassador Hotel

Panelists:

- Swati Chawla - Director of Global Operational Wellness, TP

- Ayush Verma - Senior Analyst, Resolver
- Dr. Samir Parikh - Director, Department of Mental Health and Behavioral Sciences, Fortis Mental Health
- Moderator: Uma Subramanian - Co-Founder, Rati Foundation

Key Discussions:

The panel focused on the human cost of digital safety work. Swati Chawla described how moderators are exposed to vicarious trauma and highlighted proactive interventions including mental health check-ins and a "Nice Button" allowing moderators to pause without penalty. Ayush Verma discussed rotational workforce structures to prevent burnout and training managers to recognize subtle cues like silence, irritability, or disconnection. Dr. Samir presented research showing 11% of moderation workers face measurable anxiety disorders, with symptoms including hyperarousal, numbness, and moral injury from repeated exposure to violent content. He



criticized wellness tokenism, noting abysmally low utilization rates for counseling programs, and advocated for peer mentoring and early interventions modeled on post-war trauma studies. The panel emphasized that AI can triage content but empathy cannot be automated, with humans still bearing the emotional weight of borderline cases. Uma Subramanian called for integrating "Care by Design" as a Trust & Safety benchmark rather than a wellness perk.

Takeaways:

- Trust and Safety work involves emotional labor that leads to PTSD, compassion fatigue, and moral injury.
- Surface-level wellness programs are insufficient; care must be systemic and embedded in operational DNA.
- Permission to pause, honest conversations without stigma, and peer validation systems are critical protective factors.
- The first content needing protection is the mind of the person reviewing harmful material.

Recommendations:

- Invest in wellness-focused innovation within Trust & Safety operations.
- Embed Care-by-Design principles as practical benchmarks for healthier platform practices.
- Shape industry norms for rotational duties, peer support, and early trauma intervention.
- Build non-stigmatizing support spaces for moderators experiencing occupational stress.

3.14. Panel: Strong & Seen: How Women Are Leading Learning, Enterprise, and Solidarity Online with YouTube

Date: October 8, 2025

Time: 4:45 PM

Venue: Dome, Taj Ambassador, New Delhi

Speakers:

- Iqra Choudhary - Member of Parliament, Kairana, Uttar Pradesh



- Laxmi - Co-Producer, Khabar Lahariya
- Dr. Cuterus (Dr. Tanaya Narendra) - Content Creator & Educator, YouTube
- Neha Agrawal - Founder, Mathematically Inclined
- Moderator: Aparajita Bharti - Founding Partner, The Quantum Hub

Key Discussions:

The conversation opened with Laxmi of Khabar Lahariya, who described how rural women journalists use digital tools to challenge entrenched power structures and assert their right to report. She detailed how Khabar Lahariya trains women in digital literacy from operating smartphones to securing their content, so they can document stories that often remain invisible. Laxmi spoke about the social stigma women face for simply using mobile phones and highlighted how documentation through digital platforms becomes an act of collective resistance, enabling women in rural communities to reclaim their narratives.

Drawing from her own activism and political journey, Iqra Choudhary emphasized how online platforms have become essential spaces for young women to voice grievances and demand accountability on issues such as girls' education. She highlighted the emotional and political impact of gendered harassment, disinformation, and targeted takedowns, noting that safety concerns remain a barrier to participation. Iqra stressed the need for stronger digital infrastructure, legal protections, and platform-level responsibilities so that women, especially in rural and low-income communities, can speak without fear. Online solidarity, she argued, has the power to transform individual stories into collective movements.

Shifting the focus to education and health, Neha Agrawal and Dr. Tanaya Narendra explored how women educators are expanding access to learning and factual information through YouTube. Neha discussed how she built an accessible math-learning ecosystem for students who lack traditional classroom support, while also tackling persistent stereotypes about women in STEM. She emphasized sustainability and depth in teaching over viral content.

Dr. Tanaya highlighted the widespread misinformation surrounding women's health and described how her work demystifies reproductive health, HPV vaccination, and anatomy using clear, science-based communication. She stressed that openness online helps dismantle shame and provides young people, especially young women, with the knowledge they are often denied offline. Collectively, the panelists illustrated the potential of safe and accessible digital platforms to foster solidarity, expand learning, and support social transformation.

Takeaways

- Digital platforms offer women powerful ways to document reality, build community, and hold institutions accountable.
- Online risks, harassment, misinformation, and suppression, remain significant, making digital safety central to women's participation.
- Women educators and creators are leveraging YouTube to democratize learning and challenge gender stereotypes in STEM, health, and civic life.
- Solidarity online grows when women have the skills, confidence, and protection to share their stories.

Recommendations:

- Evolve policy and platform approaches to better address gendered harassment and suppression of women's voices.
- Grow digital literacy and safety programs for women through government, industry, community collaboration.
- Support women-led educational content with funding, partnerships, and platform visibility.
- Refine reporting pathways and collaborative safeguards to better address gendered online abuse.

At Google, we aim to balance delivering information with protecting users and society and our goal is to provide access to trustworthy information and content by protecting users from harm, delivering reliable information and partnering with experts and organisations to create a safer Internet. Delighted to participate at TASI as part of efforts to share and exchange knowledge to keep users safe.

Norman Ng
Head, Trust & Safety Global Engagements (APAC), Google



4. IMMERSIVE SESSIONS

4.1. Tackling Child Sexual Abuse Online – A Global Perspective (Day I)

Date: October 7, 2025

Time: 9:00 AM

Venue: Tango, Taj Ambassador Hotel

Presenters:

- Siddharth Pillai - Co-Founder, RATI Foundation
- Professor Deborah Fry - Global Director of Data, Childlight Global Child Safety Institute; Professor of International Child Protection Research, University of Edinburgh

Key Discussions:

The session brought together leading voices from India and the global research community to examine the evolving challenge of Online Child Sexual Abuse (OCSA) and exploitation. Siddharth Pillai discussed how RATI Foundation's early work on digital safety catalyzed policy and institutional responses to OCSA in India. He traced the creation of India's first citizen reporting portal for Child Sexual Abuse Material (CSAM) in partnership with the Internet Watch Foundation (IWF) a model that informed the national reporting system cybercrime.gov.in. Pillai introduced Meri Trust Line (6363-176363), a dedicated helpline launched in 2023 to assist victims of image-based abuse and other online harms, which



has facilitated over 2,000 calls and 1,500 content takedowns. He emphasized that while more victims are coming forward, re-uploading and persistence of CSAM remain major concerns.

Professor Deborah Fry expanded the discussion to global scale, presenting findings from the forthcoming Into the Light Index 2025. Focusing on South Asia and Western Europe, she revealed that one in eight children in South Asia experiences rape or sexual assault in childhood—translating to around 54 million children across India, Nepal, and Sri Lanka. Fry highlighted the widening gap between children's lived realities online and available data quality, noting a 1000% global increase in AI-generated CSAM between 2023 and 2024. The Index consolidates data from major organizations including IWF, NCMEC, and the Canadian Centre for Child Protection. She emphasized the need for open, transparent, and accessible data to guide interventions, with the Index's upcoming interactive dashboard helping researchers, civil society, and governments track emerging threats. Fry concluded by underscoring that "digital protection cannot exist in isolation from the environments that shape a child's life.

Takeaways:

- Online child sexual abuse is evolving with AI, deepfakes, and cross-platform exploitation
- Survivor-centered, trauma-informed approaches including dedicated helplines are crucial

- Data gaps remain a major barrier; collaborative, transparent data-sharing is needed
- Law enforcement, platforms, and governments must coordinate more closely for timely, sensitive case handling
- Child protection online must be embedded within broader education, welfare, and safety systems

Recommendations:

- Strengthen national helpline systems to ensure timely, coordinated support for children and families.
- Enable safer, anonymized CSAM trend-data sharing through robust national collaboration frameworks.
- Improve platform practices for CSAM detection and removal through proactive safeguards and clearer response processes.
- Embed digital safety, consent, and citizenship concepts into school curricula to build early awareness and protection.
- Center survivor and practitioner insights in policy design to ensure responses are grounded in real needs and experiences.

Tackling Child Sexual Abuse Online – A Global Perspective (Day II)

Date: October 8, 2025

Time: 3 PM

Venue: Tango, Taj Ambassador Hotel

Presenter: Abby Roberts - Project Manager, INHOPE

Key Discussions:

Building on the earlier session "Tackling Child Sexual Abuse Online – A Global Perspective," Abby Roberts presented INHOPE's Research Schema as the first globally agreed-upon data language for identifying, classifying, and reporting child-sexual-abuse and exploitation material (CSAM). The core challenge is that each jurisdiction and industry partner has historically maintained its own vocabulary, different severity ratings, metadata categories, and procedural fields, which fragments intelligence and delays victim rescue. "We cannot scale protection if we do not first speak the same language," Roberts emphasized.

The new schema proposes a shared syntax and ontology mapping every report, image, or case into a standard digital record readable by all accredited partners. Co-developed with more than forty stakeholders including hotline analysts, law-enforcement officers, child-protection NGOs, and major technology companies, its architecture allows national systems to retain local privacy safeguards while exchanging anonymized case identifiers through an interoperable API.

Roberts emphasized trust through transparency, noting that harmonized data will enable researchers to produce global prevalence estimates and identify emerging abuse typologies without exposing sensitive imagery. INHOPE's research wing intends to publish aggregated insights giving policymakers evidence that is both statistically credible and victim-centered. The discussion acknowledged that standardization should not homogenize context, developing nations may face infrastructural limits or legal differences requiring adaptive implementation through regional training and phased adoption.

Emerging Themes:

- Global harmonization of CSAM classification enables cross-border cooperation and intelligence sharing
- Interoperability between law enforcement, NGOs, and industry requires common data language
- Evidence-based reporting standards support both operational efficiency and victim protection
- Data ethics and privacy must be maintained in transnational information-sharing



4.2. Bodies, Bans & Algorithms: Reclaiming SRHR Access in the Age of Digital Censorship and Control

Date: October 7, 2025

Time: 11:00 AM

Venue: Tango, Taj Ambassador Hotel

Panelists:

- Kritika Yadav - Founder, Love Matters India
- Aparajita Bharti - Co-Founder, The Quantum Hub
- Sachee Malhotra - Co-Founder, That Sassy Thing
- Ms. Zoya Ali - Founder, Uteropedia
- Moderator - Prakshi Saha - Founder, Frida Health

Key Discussions:

The session examined how digital censorship and algorithmic bias routinely suppress SRHR content. Moderator Prakshi Saha opened by noting that platforms often mistake education for obscenity, creating safety systems that silence, rather than support, essential information.

Kritika Yadav highlighted repeated flagging of posts on menstruation and contraception, while Sachee Malhotra described how women-led wellness brands face shadow bans and unclear ad restrictions. Aparajita Bharti stressed that India's digital governance still mirrors social stigma, urging evidence-based moderation guidelines shaped with health and gender experts.

From a grassroots lens, Zoya Ali emphasized that many young people rely on social platforms for SRHR knowledge, making visibility crucial. Panelists collectively argued that real safety requires context-sensitive moderation, transparent appeals, and policies that protect educators rather than penalize them, ensuring SRHR content remains accessible, accurate, and empowering.

Takeaways:

- Automated moderation may disproportionately censor SRHR educators and women-led initiatives



- Context-sensitive safety design requires involvement of health experts and feminist advocates
- Transparent appeal and review mechanisms build creator trust
- Digital safety must coexist with the right to information and bodily autonomy

Recommendations:

- Strengthen contextual SRHR training for trust and safety teams across platforms.
- Form co-regulatory working groups that connect public health, gender justice, and tech stakeholders.
- Create an independent appeals pathway for health and education content that may be incorrectly flagged.

4.3. Decoding TFGBV: Building Policy and Governance Solutions

Date: October 7, 2025

Time: 3:00 PM

Venue: Tango, Taj Ambassador Hotel

Panelists:

- Gokul Narayan - Chief Executive Officer, Asian School of Cyber Law
- Shohini Banerjee - Knowledge Specialist, Point of View
- Kundan Mishra - Senior Program Officer, International Development Research Centre
- Vrinda Bhandari - Advocate, Supreme Court of India
- Sophie Mortimer - Manager, UK Revenge Porn Helpline, South West Grid for Learning
- Moderated by: Poulomi Pal - Program Specialist, UN Women India

Key Discussions:

Framed ahead of the AI Impact Summit, the session surfaced an uncomfortable truth about how women and LGBTQ+ people bear disproportionate harms online. Shohini Banerjee mapped TFGBV's spectrum from harassment and trolling to revenge pornography, doxxing, and covert offline-to-online abuses such as attaching tracking apps



to partners' phones. She underlined how AI has scaled harms and how platform rules struggle to offer timely relief, insisting on a survivor-centric approach with tailored pathways that prioritize agency.

Vrinda Bhandari located the problem in statutory blind spots, noting that while obvious crimes are covered, morphing, cyberbullying, and coordinated troll armies "do not fall neatly" under any single offense. She urged vernacular support and training for magistrates and local law enforcement, stronger collaboration between government and intermediaries, and greater due-diligence obligations.

From a regulatory-tech angle, Gokul Narayan argued that tech develops fast, law is a slow-moving elephant, a mismatch exploited by the pornography ecosystem. He pressed for clarity of legislative purpose and noted that framing AI chiefly as a privacy threat can be elitist when millions face immediate TFGBV harms. Sophie Mortimer unpacked the tooling, explaining how perceptual hashes help with images and advocating for StopNCII hash adoption with "little downside" and significant upside for global social good.

Kundan Mishra stressed measurement for policy traction, sharing that in an 18-country study with 20,000+ responses, 60% reported experiencing TFGBV, 40% did not seek solutions, and 25% of LGBTQ respondents who experienced TFGBV were more likely to contemplate self-harm. These numbers highlight under-reporting and distress pathways, sharpening the case for context-aware redressal.

Takeaways:

- TFGBV spans a wide continuum; AI has

scaled harms while platform takedowns seldom deliver timely relief.

- Legal gaps persist for morphing, cyberbullying, and troll armies; improve implementation and protect survivor privacy
- Technical levers including similarity-based AI, perceptual/video hashing, and StopNCII adoption can curb circulation
- Build capacity through vernacular training, local law enforcement sensitization, and co-regulation with intermediaries

Recommendations:

- Strengthen due-diligence practices and expand similarity-matching tools for TFGBV content, including broader use of solutions like StopNCII.
- Build vernacular capacity with ongoing training for magistrates, local police, and frontline responders.
- Use demographic-aware TFGBV monitoring and publish annual progress dashboards.
- Recognize targeted harassment of marginalized women in power and expand TFGBV frameworks and resources accordingly.

4.4. India in Trust & Safety: Where Digital Harms Meet Global Solutions

Date: October 8, 2025

Time: 9:15 AM

Venue: Tango, Taj Ambassador Hotel

Speaker: Dhruv Khosla - Practice Director, Everest Group

Key Discussions:

Dhruv Khosla highlighted India's unique dual



role in Trust & Safety as both a demanding market for solutions and a global hub for delivery. The session focused on factors driving T&S demand from India, contrasted Indian regulations against global frameworks, and examined T&S delivery capabilities from India including future-proofing strategies. Khosla emphasized that India's position offers unprecedented opportunities to shape both consumption and production of trust and safety services globally.

Takeaways:

- India serves as both major consumer and provider of trust and safety services
- Understanding India-specific regulatory landscape is crucial for global T&S operations
- Future-proofing requires investment in skilled workforce and infrastructure

Recommendations:

- Support development of T&S workforce capacity and training programs in India
- Align Indian regulations with global standards while respecting local context

- Foster innovation hubs for T&S technology and service development

4.5. Children's Safety By Design: Policy, Practice And Future

Date: October 8, 2025

Time: 11:15 AM

Venue: Tango, Taj Ambassador Hotel

Panelists:

- Dr. Rabindra Narayan Behera, MP, Lok Sabha and Member of Parliamentary Committee on IT
- Ms. Chitra Iyer, CEO and Co-founder, Space2Grow
- Mr. Shantaram Jonnalagadda, Director, Yoti
- Sadhna Singh, Consultant, NITI Aayog
- Moderator: Ms. Heena Goswami - Institute for Governance, Policies & Politics (IGPP)

Key Discussions:

The discussion opened with the moderator positioning Safety by Design as an urgent priority for India's child online safety ecosystem

Dr. Behera outlined the rapidly evolving threat landscape and shared how institutions are working to stay responsive amid rising cybercrimes involving children.

Sadhna Singh added a governance perspective, pointing to structural and institutional gaps that impede effective prevention and response.

Shantaram Jonnalagadda spoke to emerging safety-by-design practices, such as privacy-preserving age assurance and risk-mitigation features, being integrated across digital systems.

Adding a behavioural lens, Chitra Iyer highlighted concerns around children forming emotional dependencies on AI systems and underscored the need for holistic digital literacy and parenting frameworks.

The session was highly interactive, with audience questions foregrounding institutional accountability and systemic safety, and the panel underscored that India needs contextual

solutions and that Safety by Design must be embedded across law, education, and platform architecture.

Takeaways:

- Safety by Design must move from concept to implementation across governance and platforms.
- Capacity-building for parents, schools, and enforcement is essential for prevention.
- AI introduces new risks, deepfakes, impersonation, emotional manipulation requiring cross-sector vigilance.

Recommendations:

- Expand India-specific research on emerging child safety risks, including digital self-harm.
- Embed safety-by-design principles into product development and ecosystem practices.
- Strengthen gender-sensitive child protection measures across platforms.
- Scale age-appropriate digital literacy programs for children and caregivers.



5. WORKSHOPS

5.1. Designing Safer Pathways: Introducing GroSafe's AI Chatbot for Child Protection

Date: October 7, 2025

Time: 12:00 PM

Venue: Tango, Taj Ambassador Hotel

Facilitator: Dylan Schouten - Research Project Manager at GroSafe at TU Dublin

Workshop Highlights:

This interactive workshop introduced GroSafe, an EU-funded initiative using gaming, knowledge management, and AI to protect children from online grooming. Participants

explored GroSafe's in-development AI reporting chatbot designed to make it easier for young people to seek help and report concerns. Mr. Schouten emphasized that the GroSafe tool differs from traditional mass-consumer agentic AI, it does not mimic users' speech tones, attempt to befriend them, or confirm biases. Operating from data in a closed server, the GroSafe AI tool only serves as a signposting agent, directing users to appropriate helplines when it detects signs or signals of grooming or abuse.

Through breakout discussions and hands-on testing, attendees provided feedback on the chatbot's design, usability, and implementation challenges. The group debated the balance between speed and



sensitivity, with technologists cautioning that false positives could overwhelm support lines while mental-health professionals emphasized the value of erring on the side of care.

Key Outcomes:

- Ethical AI can support early intervention in mental-health crises and grooming situations without replacing human care networks
- Design must prioritize trust over surveillance, "People seek help when they feel safe, not when they feel watched"
- Continuous feedback loops and human review maintain trust and effectiveness

Action Points:

- Support public-interest research on AI for mental health and safety applications through grants and regulatory sandboxes

- Develop national guidelines on ethical AI-mediated well-being tools emphasizing consent and accountability
- Integrate community feedback into ongoing GroSafe development and implementation

5.2. Strategic Foresight: Future-Proofing Trust

Date: October 7, 2025

Time: 01:30 PM

Venue: Tango, Taj Ambassador Hotel

Facilitator: Farshida Zafar - Director, Erasmus Centre for Entrepreneurship and Jip Dresia

Workshop Highlights:

The workshop invited participants to step beyond prediction into preparation, positioning strategic foresight as essential methodology for navigating volatile digital landscapes.



Farshida Zafar framed Trust & Safety not merely as a corporate function but as a discipline with its own ecosystem of actors: policymakers, researchers, technologists, and communities.

Facilitators outlined a six-step foresight methodology: (1) Unpack trends through PESTEL analysis, (2) Filter relevance, (3) Focus on key uncertainties, (4) Craft multiple scenarios combining trends and uncertainties, (5) Forecast scenario developments, (6) Reverse-engineer responses. The workshop stressed that outliers, not trends, most often destabilize trust-trends are slow and predictable while outliers are abrupt and high-impact.

Participants engaged in scenario exercises exploring futures for 2035, examining key uncertainties such as whether the world would move toward centralized governance or decentralized protocols, and whether technology would evolve toward human-aligned or autonomous systems. Scenarios ranged from a "Global Digital Accord" with extensive surveillance to an "AI Wild West" with fragmented trust, to a "Global Trust Commons" requiring constant cross-sector dialogue.

Key Outcomes:

- Foresight outpaces forecasting - planning beyond immediate horizons enables organizations to manage disruption proactively
- Uncertainties, not trends, define risk; outliers like generative AI can fracture trust
- Regulation vs autonomy presents core trade-offs; over-regulation risks conformity while under-regulation risks chaos
- Trust requires multi-stakeholder participation across governments, industry, academia, and citizens

Action Points:

- Institutionalize foresight units within digital-policy and Trust & Safety departments to conduct recurring scenario analyses
- Integrate foresight training into corporate and academic programs
- Develop early-warning indicators for trust erosion, linking M&E systems to real-time policy adjustment
- Encourage cross-border foresight collaborations between India, ASEAN, and EU partners



5.3. Care, Not Control: An Intergenerational Dialogue on Teens' Online Safety

Date: October 7, 2025

Time: 4:15 PM

Venue: Tango, Taj Ambassador Hotel

Facilitators:

- Aparajita Bharti - Co-Founder YLAC & The Quantum Hub)
- Kevin Lee - CEO, Yuva
- Sagar Deoskar - Lead, Public Policy for South Asia, Snap Inc.

Workshop Highlights:

This intergenerational dialogue brought together teens, parents, educators, and civil society to explore the tension between trust and safety in digital parenting. Deepak Goel(DDG at the Cyber Laws Division, MEITY) framed the central tension: "As a child, trust comes first; as a parent, safety comes first." The facilitators established ground rules including non-judgment, openness, inclusivity, and "Vegas Rule" confidentiality.

The workshop's format encouraged open, non-judgmental sharing under "Vegas Rule" confidentiality, placing teens' voices at the center. Teens described care as a supportive "safety net" and control as pressure, distinguishing concern after behavioral change from stifling pre-emptive scrutiny. Parents spoke of anxiety, fear of empathy erosion, and limited digital literacy, while educators and civil-society representatives pointed to isolation, attention loss, and weak reporting mechanisms.

Rapid-fire exchanges revealed divergent priorities: teens value emotional safety and agency; parents focus on exposure risks; educators lament attention decline; civil-society advocates emphasize structural gaps. When asked about fair parental rules, teens highlighted time limits, transparency, and open dialogue.

The workshop concluded on the principle that digital safety must be co-created. Parents need digital empathy, children must see boundaries as protection, and institutions play a role in bridging trust and safety, captured in the metaphor "care as a net, not a cage."



Outcomes

- Digital well-being relies on empathy and open dialogue, not punitive monitoring.
- Co-regulation works better than control, and shared boundary-setting is what ultimately builds safer, trust-based digital ecosystems.

Action Points:

- Build parental controls as transparent, co-use tools and provide plain-language, vernacular guidance for parents and educators.
- Shift from punitive age-gating to co-regulation, with safety standards co-created by youth, platforms, and policymakers.

5.4. Reimagining Safety: Youth, Tech, and Active Citizenship

Date: October 8, 2025

Time: 12:15 PM

Venue: Tango, Taj Ambassador Hotel

Facilitator: Sonali Vyas - Director, Program & Operations, Safetipin

Workshop Highlights:

The workshop invited participants to rethink safety as a participatory, citizen-led process rather than a top-down mandate, using Safetipin's interactive, crowd-sourced mapping tools to show how real-time data can shape safer cities. Facilitators Sonali Vyas and Tanya Thakur introduced the app's philosophy of "safety through participation" before guiding a thought experiment where groups stepped into the daily lives of a young commuter, a middle-aged professional, and an elderly teacher to map their routes and identify context-specific risks. As participants visualized these journeys, they saw how safety varies by gender, age, ability, and mobility, and how user-generated data has already influenced better lighting, redesigned bus stops, and safer walking routes. The discussion also connected physical and online safety, noting that the same technologies that aid navigation can introduce digital risks. The session closed with a reflection on active citizenship and the power of youth-led audits and data-driven advocacy to make cities more inclusive for those whose safety concerns are often invisible.



Key Outcomes:

- Youth-led safety innovations prioritize community agency over institutional control
- Digital tools can bridge online and offline safety when designed with marginalized communities
- Active citizenship emerges when young people have platforms to document, share, and advocate

Action Points:

- Support youth-led safety auditing and mapping initiatives with funding and technical resources
- Integrate community-generated safety data into policy and urban planning
- Create platforms for peer-to-peer safety networks that center marginalized voices

5.5. Sandboxing a Regulatory Sandbox for Child Online Safety

Date: October 8, 2025

Time: 2:00 PM

Venue: Tango, Taj Ambassador Hotel

Format: Workshop with Panel Discussion

Facilitator: Dr. Ioanna Noula - Founder, The Children's Online Redress Sandbox

Panelists:

- Bhumika Chaturvedi - NLU Jodhpur
- Prof. (Dr.) Shilohu Rao - Professor (Law and Tech) & Director, Centre for Law and Technology (NLU Jodhpur)
- Vaishnavi J - Founder, Vys
- Adv. Ghanshyam Meena - NLU Jodhpur

Workshop Highlights:

This hybrid workshop-panel introduced the COR Sandbox as an experimental regulatory space where youth, academics, technologists, and lawyers co-test policy ideas and complaint systems before they solidify into law. Dr. Ioanna Noula underscored that ethical innovation demands children's direct involvement, while Prof. Rao highlighted the India-Greece partnership as a way to blend cultural context with strong privacy norms and let policy "learn faster than harm."

Adv. Ghanshyam Meena stressed that grievance mechanisms must be simple and child-friendly, not bureaucratic.



Vaishnavi J explained that COR goes beyond token youth consultation by embedding child feedback into product design and policy review, supported by academic-industry collaborations that help smaller companies strengthen safety design.

Key Outcomes:

- Participation is protection; COR redefines safety by co-designing systems with children, giving agency a central role
- Sandboxes accelerate learning, allowing law and policy to evolve faster than harm
- Youth feedback must move from consultation to co-creation reflected in final design
- Academic-industry partnerships fill capacity gaps for safety design infrastructure

Action Points:

- Set up child-centered sandboxes in universities where minors can try out safety mechanisms before wider rollout.
- Integrate co-design into governance processes by building in measurable

youth-participation components.

- Design online redressal portals for child complaints with clear, plain-language interfaces.
- Strengthen cross-border academic collaboration to co-develop evidence-based safety standards.

5.6. #CreateWithCare: A Roundtable on India's Content Ecosystem & Responsible Practices

Date: October 8, 2025

Time: 4:00 PM

Venue: Tango, Taj Ambassador Hotel

Format: Closed-Door Roundtable

Organized by: ACTS (Alliance for Cyber Trust and Safety)

Participants:

30+ creators, industry voices, experts, and educators including digital content creators, wellness educators, youth advocates, representatives from ASCI India, platform policy leaders, and online gaming industry stakeholders



Digital educators raised the issue of "wellness content without wellness training," describing how many influencers offer personal-health and mental-health advice despite lacking professional qualifications. "Followers mistake visibility for expertise," participants noted, proposing clear disclaimers, platform-level tags distinguishing opinion from certified information, and collaborations with licensed experts.

Youth-wellness advocates argued that the problem extends to algorithmic logic of virality, "What spreads fastest isn't what's safest." They urged platforms to design ethical friction, such as prompts encouraging fact-checking or rewarding credible sourcing rather than sheer frequency.

Content creators addressed India's weak disclosure culture, noting many do not announce when posts are paid or sponsored despite ASCI guidelines. They called for shared accountability between platforms and regulators: "If algorithms amplify content, they must also amplify transparency." Participants emphasized that audience trust, once broken, is difficult to rebuild.

The ASCI India representative outlined the Council's evolving role from enforcement to education through new certification modules helping creators identify risky language in advertising particularly for health, finance, and child-related content. "We don't want to police creativity, but to ensure creators don't mislead, even unintentionally."

Platform policy experts offered international perspectives grounded in global experience, reflecting that reactive moderation often follows public backlash but proactive responsibility must be embedded at design

stage. India's creator economy now faces the same accountability questions that shaped earlier phases of global social-media policy, how to balance open expression with societal safety.

A representative from India's online-gaming sector shared a reflective analogy. The industry, long seen as a driver of digital innovation, faced challenges when a few harmful apps affected minors and vulnerable users. Subsequent government action, they noted, highlighted the value of early, proactive self-governance. They encouraged creators to reflect on this experience and work together to strengthen safety practices, emphasising that trust is far easier to build when sectors act collectively and ahead of regulation.

The moderator concluded that Indian creators are entering a maturity phase where professionalism, literacy, and empathy must define success as much as reach. Over-regulation risks silencing creativity, but absence of standards risks exploitation. "Care isn't censorship; it's competence with conscience."

Emerging Themes:

- Ethical responsibility in India's creator economy requires balancing freedom, influence, and accountability
- Professional accountability is needed for health and mental-health content creation
- Transparency in sponsorship and partnerships remains weak despite ASCI guidelines
- Shaping algorithms to highlight credible content can contribute to a more trustworthy digital environment.
- Proactive sector-wide standards can strengthen trust and minimise the need

for external regulation.

Suggested Follow-ups:

- Develop creator certification programs on ethical content practices and disclosure requirements.
- Strengthen enforcement of transparency requirements with platform-level disclosure tools.
- Create industry-wide standards for health, wellness, and child-related content.
- Establish feedback loops between creators, platforms, and regulators for continuous improvement.
- Draw on lessons from the gaming sector to strengthen trust within the creator economy.



“When brilliant minds converge with one mission, extraordinary things happen. T&S Festival India 2025 is where visionaries and changemakers meet to challenge the status quo—where AI aligns with ethics, communities are protected, and innovation serves humanity first. This isn't just a conference, it's where breakthroughs are born.

Akash Pugalia
Chief Digital Officer, Teleperformance



6. SPOTLIGHTS

Spotlight: #CareNotControl

Presented by: Uthara Ganesh, Public Policy Head – India & South Asia, Snapchat and Nikhil Taneja, Co-Founder & Chief, Yuva

An exploration of Snapchat and Yuva's Care, Not Control campaign, which encourages open, trust-based conversations between teens and caregivers through empathetic, privacy-respecting online safety tools.



Spotlight: Tackling Image-Based Abuse – Cross-Platform Collaboration with StopNCII.org

Presented by: Sophie Mortimer, Manager, Revenge Porn Helpline

An overview of StopNCII.org's global mechanism enabling survivor-led action against non-consensual image sharing.

Spotlight: Improving Online Safety for Women and Children in India

Presented by: Natasha Jog, Head of Public Policy – Meta, India and Sidharth Deb, Associate Director, The Quantum Hub

A rapid assessment of India's legislative landscape, global regulatory models, and pathways to strengthen safety for women and children.



Spotlight: GSMA Mobile Gender Gap Research

Presented by: Akansha Kasera, Senior Market Engagement Manager, GSMA

Key insights into gendered disparities in mobile access, digital skills, and participation across low- and middle-income countries.



Spotlight: Strengthening AI Integrity – From Moderation to Mitigation

Presented by: Deepesh Gupta, Senior Director, TP

A look at emerging AI integrity challenges and how platforms can move from reactive moderation to proactive mitigation.



Spotlight: Google's Approach to Tackling Fraud and Scams

Presented by: Norman Ng, Head, Trust & Safety Global Engagement (APAC), Google

A breakdown of Google's threat detection methods, fraud-prevention strategies, and user safety interventions.



Spotlight: Regulatory Sandbox – The Children's Online Redress Sandbox

Presented by: Dr. Ioanna Noula, Founder, The Children's Online Redress Sandbox

An introduction to child-centered regulatory innovation enabling safer online environments and accessible grievance mechanisms.



Spotlight: AI Companions for Kids – Safety, Suitability & Attachment Risks

Presented by: Vaishnavi J, Founder, Vys

A session on designing age-appropriate AI companions, evaluating risks of emotional over-attachment, and applying safeguards using Vys' youth-focused AI framework.



I'm honoured to support India's inaugural Trust & Safety Festival, recognizing the country's role as one of the world's largest technology markets and a global center of innovation. I look forward to seeing the world gather here to show that innovation can truly go hand in hand with the safety and wellbeing of children.

Vaishnavi J
Founder, Vyanams Strategies (Vys)

We're proud to sponsor the inaugural Trust and Safety event and excited to share how technology can empower the travel industry and our NGO partners in fostering human rights and safety around the globe.

Neda Niazian
Director Group Trust & Safety, Booking.com

7

DIPLOMATIC PARTNERS & INTERNATIONAL ENGAGEMENT

Opening Reception at the Embassy of France - Building a Safer, More Equitable Digital Future

The Trust & Safety India (TASI) Festival 2025 commenced with an exclusive opening reception hosted at the Embassy of France in New Delhi. The evening set the tone for two days of critical exchange and collaboration on advancing a safer, more equitable digital future.

Welcoming over 100 distinguished guests representing governments, diplomatic missions, technology companies, academia, and civil society, the event underscored the power of multilateral cooperation in shaping global standards for online safety and responsible AI.

Opening remarks were delivered by H.E. Mr. Thierry Mathou, French Ambassador to India; Dr. Ranjana Kumari, Director of the Centre for Social Research; Mr. Jean-Christophe (J-C) Le Toquin, Co-Founder of the Trust & Safety Festival; and Ms. Delphine O, Ambassador-at-large and Secretary-General for the Generation Equality Forum. Their addresses highlighted the urgency of inclusive governance in digital transformation, emphasizing collaboration between the Global North and South, and the centrality of gender equality and youth participation in digital policy.

The evening also featured a Global Innovators Spotlight, showcasing cutting-edge solutions driving safety, justice, and accountability in online ecosystems. Presentations included:



- GroSafe (Netherlands): AI-powered pathways for child protection (Dylan Schouten)
- Children's Online Redress (COR) Sandbox (UK): the first regulatory sandbox for child online safety (Ioanna Noulas)
- INHOPE (EU): strengthening global coordination against child sexual abuse material (Abby Roberts)
- Stop The Traffik (UK): intelligence-led disruption of trafficking networks (Nick Dale)
- Revenge Porn Helpline (UK): preventing non-consensual intimate image sharing (Sophie Mortimer)
- ACTS (India): community-driven models for inclusive digital safety (Garima Saxena)

The reception concluded with informal networking over refreshments, fostering dialogue between international innovators, policymakers, and practitioners. The evening reflected the Festival's ethos, that trust and safety in the digital age must be co-created across borders, disciplines, and lived experiences.

“

The French Embassy in India is proud to support the Trust & Safety Festival, a crucial platform fostering global cooperation on digital safety and governance. Strengthening trust in digital spaces is essential for the future, and this festival brings together key stakeholders to advance meaningful solutions.

Thierry Mathou
Ambassador of France to India

”



Evening Reception at the Embassy of Netherlands

Following an intensive day of panels, showcases, and workshops at the Trust & Safety India (TASI) Festival, delegates convened for an exclusive evening reception hosted by the Embassy of the Kingdom of the Netherlands. Themed “Safety by Design: Women Leading Tech Solutions,” the gathering created an intimate space for reflection and forward-looking dialogue on embedding gender equity in digital innovation.

The evening began with a warm welcome by H.E. Marisa Gerards, Ambassador of the Netherlands to India, who reaffirmed her country's commitment to fostering inclusive, human-centric technology ecosystems. She highlighted the importance of global cooperation and women's leadership in building safer, more equitable digital spaces.

The reception featured a fireside chat moderated by Farshida Zafar, Director, Erasmus Centre for Entrepreneurship. The discussion brought together a distinguished panel including Major General K. Narayanan (Head of Security Law and Strategic Programs, NITI Aayog), Dr. Subi Chaturvedi (SVP & Chief Corporate Affairs & Public Policy Officer, InMobi), and Mr. Abhishek Singh (Additional Secretary, Ministry of Electronics and Information Technology, Government of India). Speakers reflected on how gender-responsive policies, safety-by-design principles, and cross-sectoral partnerships can create more resilient and inclusive digital environments. The conversation underscored that women's leadership is not just an inclusion metric but a strategic imperative for innovation, governance, and ethical AI.

Set against the backdrop of the Netherlands' continued support for digital rights and gender



equality, the evening fostered candid exchanges between government representatives, tech leaders, and civil society. The reception concluded with an informal networking dinner, marking a fitting close to Day One of TASI 2025, uniting shared commitments to safety, equity, and innovation in the digital age.

“

TASI provides a timely forum for advancing the conversations that will define the future of safe and responsible AI. At OpenAI our mission is to build AI that benefits all of humanity, and events like these help inform how we approach safety, transparency and real world deployment

Pragya Misra
Lead Public Policy & Partnerships, OpenAI, India

”



Closing Reception at the Embassy of Sweden

Following two transformative days of panels, showcases, and global dialogue at the Trust & Safety India (TASI) Festival 2025, delegates gathered for a memorable Closing Reception hosted by the Embassy of Sweden in India. The evening celebrated collective achievements while charting the path toward the AI Impact Summit 2026.

The reception opened with remarks by Jyoti Vadehra, Head of Media & Communications, Centre for Social Research, who captured the spirit of TASI 2025, its breakthroughs, collaborations, and momentum, and outlined the journey ahead as the ecosystem prepares for India's landmark AI Impact Summit 2026.

H.E. Jan Thesleff, Ambassador of Sweden to India, followed with an address marked by warmth and candour. He reflected on Sweden's longstanding commitment to international

cooperation, human rights, and digital responsibility, emphasizing that global challenges in the digital age demand shared values and collective action.

Dr. Ranjana Kumari, Director, Centre for Social Research, expressed deep gratitude to the festival's partners, delegates, and participants. She underscored TASI's vital role in connecting visionaries across continents and affirmed that meaningful progress in trust and safety emerges only when governments, industry, and civil society work in concert.

A thought-provoking panel discussion, "Looking Ahead: Lessons from TASI 2025 and the Path to the AI Impact Summit 2026," followed. Moderated by Ms. Agnes Julin, Deputy Head of Mission, Embassy of Sweden, the conversation brought together industry leaders and global online safety experts. The panel explored how innovation, cross-sector collaboration, and accountability can forge



stronger, more resilient digital ecosystems in preparation for the summit.

Against the backdrop of the Embassy's gracious hospitality and Sweden's continued leadership in digital rights and responsible innovation, the evening sparked candid conversations and forged new connections among government representatives, industry leaders, and civil society advocates.

As delegates moved into an informal networking dinner, the atmosphere hummed with possibility. The conversations that began over two days in Delhi would ripple outward informing policies, shaping platforms, and strengthening the foundations of digital trust across borders. TASI 2025 concluded not with a final word, but with an invitation: to carry forward the vision, partnerships, and resolve needed to build a digital future worthy of the societies we serve. The journey to TASI 2026 had already begun.



“

Sweden is honored to be a partner to the Trust & Safety Festival. This platform gathers key stakeholders in order to address the crucial challenges and possibilities associated with digital and emerging technologies through global cooperation. We are confident that this festival will help advancing the objective of fostering an open, safe, secure and sustainable digital future, not least with regard to the upcoming AI Impact Summit 2026 – a milestone event – hosted by the Government of India.

Jan Thesleff
Ambassador of Sweden to India

”



Address by High Commissioner of Canada to India, Christopher Cooter

Canada has been a valuable source of support to TASI, strengthening its growing role in global conversations on trust, safety, and responsible technology. This support reflects a shared commitment between Canada and India to advance AI innovation grounded in ethics, safety, and human rights.

At the festival, High Commissioner Christopher Cooter delivered remarks on October 7, highlighting Canada's longstanding leadership in AI ethics and digital rights. He also emphasized the importance of international collaboration and knowledge exchange in addressing complex technological harms. His participation underscored Canada's belief that responsible AI development must be informed by diverse lived experiences and include voices from across the world, particularly from the Global South.



8. RECOMMENDATIONS

Building on TASI 2025: Shaping India's Digital Future

Civil Society As A Catalyst For Cross-sector Innovation:

TASI 2025 proved that thoughtfully designed, civil society, convened platforms can unlock unprecedented collaboration between government, industry, academia, and citizens. By creating neutral ground for open dialogue, these forums build the mutual trust essential for co-creating regulatory frameworks and industry standards that work. The path forward requires companies and regulators to deepen their investment in these collaborative spaces, not as occasional participants, but as ongoing partners in shaping India's digital governance model.

The Ecosystem is Ready For Co-creation:

TASI 2025 revealed a decisive shift: companies are ready to move beyond compliance-driven engagement and actively co-design solutions alongside government, researchers, and civil society. Participants deeply valued candid, transparent conversations with industry leaders and expressed strong appetite for more technical and operational exchanges. This signals a strategic opening for companies to position themselves as thought leaders and innovation partners, not just regulated entities. The next phase demands clearer pathways, structured forums, collaborative working groups, and co-creation mechanisms that channel this readiness into tangible outcomes.

Implementation Expertise Drives Ecosystem Maturity:

The sessions that resonated most powerfully were those where companies opened the hood, sharing actual tools, workflows, case studies, and lessons from real-world interventions. Stakeholders consistently emphasized that these practical demonstrations bridge the gap between policy ambition and operational reality. Future editions should prioritize hands-on workshops, live walkthroughs, product demos, and implementation clinics where industry partners translate regulatory expectations into operational practice. Companies that share their learnings transparently will elevate the ecosystem's collective capability while shaping how policies are implemented at scale.

Cross-industry Perspectives Unlock Holistic Solutions:

While TASI brought together platforms, NGOs, and policymakers effectively, participants identified meaningful gaps in representation from gaming, telecom, fintech, e-commerce, and emerging tech sectors. Digital safety challenges increasingly transcend traditional platform boundaries, fraud spans fintech and social media, youth safety touches gaming and streaming, misinformation flows across news and commerce. A broader constellation of industries at the table will generate more holistic, resilient solutions. Ecosystem partners from all digital sectors should join these conversations and bring their unique operational insights to bear on shared challenges.

Sustained Engagement Amplifies Impact:

TASI demonstrated that progress requires consistent and engaged year-round collaboration. Participants emphasized that trust deepens and collaboration becomes more productive when dialogue is regular, predictable, and ongoing, not concentrated in a single annual event. Establishing working groups, innovation labs, and recurring check-ins will enable companies, regulators, and civil society to jointly prototype approaches, track shared progress, and iterate on emerging challenges together. This sustained model positions all stakeholders to anticipate developments, influence standards, and deploy solutions effectively. This approach ensures that safety is not treated as a crisis response but as a continuous responsibility. It empowers every stakeholder to move from reactive fixes to proactive, long-term impact.

India-first Innovation For Global Influence:

From AI ethics to youth protection and gendered online harms, TASI participants converged on a powerful insight: India's unique cultural, linguistic, and social context demands homegrown frameworks, not imported playbooks. The Global South perspective resonated strongly, revealing tremendous demand for contextualised, region-specific safety approaches. This represents an innovation frontier, not a constraint. Companies investing in India-centric research, localized product development, and culturally grounded safety mechanisms will serve Indian users better while pioneering models that resonate across emerging markets globally. Co-investment in research initiatives, regional pilots, and knowledge-sharing networks can ground global technologies in local realities and position India as a leader in inclusive digital governance.

Localization drives relevance and efficacy:

On issues from AI ethics to youth safety and gendered harms, participants widely agreed that India must develop context-specific, culturally anchored trust and safety strategies rather than simply adapting frameworks designed for Western markets. India's linguistic diversity, social structures, digital adoption patterns, and cultural norms create unique challenges and opportunities that generic global approaches often miss. Effective solutions require grounding in local realities, understanding how harm manifests differently across regions, languages, and communities, and designing interventions that resonate with Indian users' lived experiences. This localization imperative isn't about rejecting global best practices but about ensuring that safety mechanisms, content policies, and governance models actually work in the Indian context.

AI Governance Demands Coordinated Action:

TASI 2025 surfaced the urgent need for aligned approaches to AI governance, spanning risk assessment, transparency, accountability, and societal impact. As generative AI, automated content moderation, and algorithmic systems become embedded across sectors, the gap between technical capability and governance readiness grows acute. This requires intensified collaboration: companies bringing technical depth and operational constraints, regulators contributing policy vision and public interest mandates, researchers offering evaluation frameworks, and civil society grounding discussions in user experience and rights. Dedicated AI safety working groups can jointly develop practical governance frameworks, share red-teaming insights, and pilot transparency mechanisms that work at scale.

Building On The Momentum:

TASI 2025 has demonstrated what's possible when diverse stakeholders commit to the spirit of collaboration. The ecosystem is primed for deeper technical exchanges, sustained working relationships, and India-led innovation in digital governance. Companies ready to lead with transparency, regulators prepared to co-design adaptive frameworks, and ecosystem partners across sectors can now transform this momentum into measurable impact. India has the opportunity to model participatory digital governance where innovation and safety advance together, and the pathway forward is clear.

“

This event will be a game-changer. What a privilege it will be for me to participate in the Global Majority's first Trust and Safety Festival, bringing together top digital wellbeing experts from multiple disciplines, sectors and countries. I can't wait!

Anne Collier
Founder, The Net Safety Collaborative, Strategy Lead, COR
Sandbox

”

9. AS WE LOOK FORWARD

Looking Ahead to TASI 2026

TASI 2025 marked the inaugural edition of the Trust and Safety India Festival and set an important foundation for the conversations ahead. It brought together government leaders, industry partners, civil society organisations, and academics for a rare, solutions-focused dialogue on the future of digital safety in India. Although the festival took place in India, the conversations were global in scope, reflecting the increasingly important role that India plays in shaping digital governance and safety frameworks for emerging technologies.

As we prepare for 2026, I am encouraged by the clarity and energy that emerged from this first year. The festival demonstrated that cross-sector cooperation is both possible and essential, and it created a space where difficult issues could be discussed with honesty and intent. My hope is that the next edition will broaden participation from industry, introduce new areas of focus, and deepen the quality of engagement across all stakeholders. With the India Impact AI Summit setting the stage earlier in the year, TASI 2026 in October will have the opportunity to build on those global conversations and carry them forward in a more focused and practical way.

Grateful to everyone who participated in and contributed to a successful inaugural TASI in 2025 and looking forward to an even more ambitious and inclusive TASI 2026.

Karuna Nain

Online Safety Expert and Advisor to CSR, India

As technology challenges facing our interconnected communities continue to grow in scale and complexity, this year's Trust and Safety Festival taking place in India's capital spotlights the urgent need for meaningful, multistakeholder, and decentralised conversations on online trust and safety that account for the untapped strengths, and under-researched vulnerabilities of online communities across the Majority World

Dr Ioanna Noura
Founder and Project Lead, COR Sandbox

Building safer digital spaces for girls requires more than safeguards - it requires trust, inclusion, and shared accountability. The India Trust & Safety Festival is a timely reminder that when girls are empowered online, we strengthen not just their futures but the fabric of our society.

Major Sadhna Singh (Retd)
Consultant, Home & Defence, NITI AAYOG

10. THE NEXT CHAPTER - TASI 2026

As we look ahead, TASI's next chapter is guided by a simple but powerful imperative: to deepen its role as a convening space where diverse voices shape India's digital safety future collectively and meaningfully. This year has shown us that when practitioners, communities, policymakers, industry leaders, and young people come together, a different kind of conversation becomes possible, one grounded in lived experience as much as in technical expertise. Building on this momentum, TASI 2026 aims to grow into an even more inclusive, grounded, and globally resonant platform that not only hosts dialogue, but actively strengthens the ecosystem's ability to imagine, implement, and measure safer digital futures for all.

A Stronger Global South Voice Shaping Digital Futures:

TASI 2026 envisions a deeper consolidation of its identity as a Global South-led space where emerging economies define their own priorities for digital safety, governance, and rights. By bringing together countries with similar socio-technical contexts, TASI seeks to build a shared vocabulary, advance South-South learning, and elevate perspectives often underrepresented in global conversations.

A More Interconnected Ecosystem With Broader Industry Participation

As digital interactions expand across sectors, so do the safety challenges that accompany them. TASI 2026 aspires to welcome richer

participation from telecommunications, gaming, fintech, edtech, ecommerce, media, and other adjacent industries. Their inclusion can deepen cross-sector understanding and help shape solutions that reflect the realities of a rapidly evolving digital landscape.

A Culture of Accountability and Measurable Progress

One of the festival's future ambitions is to introduce a Trust & Safety Index that captures the ecosystem's year-on-year progress. Such a benchmark would help readers, partners, and practitioners understand where advancements are being made, across AI governance, gendered harm mitigation, crisis management, and institutional capacity, and where collective attention is still needed.

A More Inclusive and Regionally Grounded Platform:

Looking ahead, TASI hopes to strengthen engagement with young people and communities beyond India's major cities. State-level dialogues, youth labs, and partnerships with schools and universities can create new entry points for diverse voices, ensuring that future digital safety conversations are responsive to varying geographies, social contexts, and lived experiences.

Collaborative Spaces That Turn Ideas Into Practice

To support deeper problem-solving across the ecosystem, future editions may introduce

formats such as solution-focused roundtables and hands-on implementation clinics. These spaces would allow policymakers, researchers, practitioners, and industry teams to collectively explore emerging questions, from age assurance and crisis escalation to transparency reporting and AI governance, with a focus on translating insight into action.

A Hub for Shared Learning, Experimentation, and Innovation

TASI 2026 also imagines a stronger emphasis on practical knowledge exchange. Case walk-throughs, reflections on failures, and innovation showcases featuring safety-tech pioneers and research labs can create opportunities for readers and participants to engage with real-world tools, methods, and lessons, supporting a culture of evidence-led improvement and collaborative problem-solving across sectors.

Cross-Cutting Principles

These recommendations are grounded in several shared commitments:

- **Safety cannot be an afterthought:** It must be embedded in design, policy, and practice from the outset
- **Multistakeholder collaboration is essential:** No single actor can address these challenges alone
- **Global South leadership matters:** Solutions must reflect diverse contexts and be shaped by those most affected

- **Transparency enables accountability:** Sunlight remains the best disinfectant for digital harms
- **Care is foundational:** For users, for professionals, for communities navigating digital spaces

The path from recommendations to implementation requires sustained commitment, adequate resourcing, and genuine partnership across sectors. TASI commits to tracking progress on these recommendations and providing platforms for continued dialogue as we work collectively toward safer digital futures.

NOTE OF GRATITUDE

After years of working at the intersection of gender justice, and digital wellbeing, we have witnessed firsthand how conversations about online safety often happen in silos - tech companies speaking to tech companies, policymakers crafting frameworks in isolation, civil society advocating from the margins, and most critically, those directly affected by digital harms rarely finding space at decision-making tables.

TASI 2025 was born from a conviction that this fragmentation serves no one, least of all the women, children, and marginalized communities bearing the brunt of online harms. We believed that India and the broader Global South needed a platform where these conversations could happen not in parallel, but in genuine dialogue. Where the voices of survivors would not be tokenized but centered. Where policymakers, platforms, researchers, and advocates could engage not as adversaries but as collaborators working toward shared goals.

What unfolded over October 7-8, 2025 exceeded even our most optimistic hopes. To bring together over 300 participants from 15 countries, to host ministerial-level keynotes alongside grassroots advocates, to facilitate substantive dialogue across six complex tracks.

This was made possible by extraordinary partnerships. Thirty-two organisations across technology platforms, civil society, government missions, academia, and specialised initiatives came together to shape TASI 2025. Their collective presence affirmed that India has long been waiting for a platform like TASI, one where voices from the ground finally have a seat at the table and every stakeholder in the digital safety ecosystem is represented.

Jyoti Vadehra

Head, Online Safety and Wellbeing, CSR India

SPECIAL THANKS

To Dr. S. Jaishankar, External Affairs Minister, for delivering a keynote address that elevated the entire festival and affirmed the critical importance of trust, accountability, and inclusion in AI governance. Your participation sent a powerful signal about the government's commitment to these issues.

To S. Krishnan and Abhishek Singh from the Ministry of Electronics and Information Technology, for your thoughtful engagement, candid conversations, and willingness to sit in dialogue with diverse stakeholders. Your presence demonstrated that government is listening, and that matters deeply.

To our international partners, the Embassy of France, Embassy of the Kingdom of the Netherlands, Embassy of Sweden, Canadian High Commission, and Australian High Commission, thank you for recognizing that trust and safety is a global challenge requiring international solidarity. Your support helped position India at the center of this crucial conversation and affirmed TASI's role as an official pre-summit event for the AI Impact Summit 2026.

To the technology platforms and companies, Booking.com, Teleperformance, Meta, Snapchat, YouTube, Truecaller, GSMA, Match Group, X, Roblox, and Resolver, thank you for showing up, for engaging honestly even when conversations were difficult, and for your willingness to be held accountable while also sharing the challenges you face. Genuine partnership requires this kind of openness and your presence demonstrates commitment to moving beyond rhetoric to real dialogue.

To our civil society and research partners, GirlEffect, UN Women, The Asia Foundation, Alliance for Cyber Trust and Safety (ACTS), Safetipin, The Dialogue, INHOPE, Children's Online Redress Sandbox (COR), GroSafe, and FRIDA, you are the backbone of this work. Your research, advocacy, and direct support for affected communities grounds everything we do in lived reality rather than abstraction. You ensure that survivors' voices and community needs remain at the center.

To our academic, policy, knowledge, and mental health partners, Dhirubhai Ambani University, Institute for Governance, Policies and Politics (IGPP), The Quantum Hub, Advertising Council of India (ASCI), Obhan & Associates, Vyanams Strategies, and Fortis Healthcare, thank you for bringing rigorous research, legal expertise, policy analysis, strategic thinking, and critical mental health insights to the festival. Your contributions enriched discussions and helped translate ideas into actionable, holistic frameworks.

To our 60+ speakers and moderators, who prepared thoughtfully, spoke courageously, and facilitated conversations that were both rigorous and compassionate. You brought depth, nuance, and humanity to complex technical and policy discussions.

To every participant who traveled to New Delhi, who engaged in panels and workshops, who asked difficult questions, who shared expertise and experience - you are the TASI community. Your active participation transformed what could have been a series of presentations into genuine dialogue and collective problem-solving.

To the survivors and advocates who shared painful experiences and hard-won insights, often at personal cost, your courage and expertise are the foundation upon which all of this work rests. We are committed to ensuring that your voices continue to shape not just conversations but concrete policy and practice.

And finally, to everyone who believes that a safer, more equitable digital world is possible, thank you for choosing to build it with us. TASI 2025 was just the beginning.

At Meta, creating avenues for open dialogue and awareness is part of our long-term commitment to user safety and building platforms that prioritize people's well-being. Over the years, we've launched more than 50 safety tools and features, investing over \$20 billion in technology to promote online safety. We are proud to support the Trust & Safety Festival in India which is taking another step forward in our endeavor to create a safer and inclusive internet.

Natasha Jog
Director, Public Policy India, Meta

SUGGESTED READING & MATERIALS

- Safe & Trusted AI – IndiaAI Safety Institute: Comprehensive resources on AI safety policy, governance, and applied toolkits in the Indian context. (indiaai.gov.in/hub/safe-trusted-ai)
- Principles for Responsible AI – NITI Aayog: India's foundational paper on responsible AI development and deployment, widely used by policymakers and practitioners. (niti.gov.in)
- The Developer's Playbook for Responsible AI in India – NASSCOM: A practical guide covering ethics, transparency, and safety considerations for AI development in India. (nasscom.in)
- AI Trust and Safety Re-imagination Programme – UNDP: Global proposals and frameworks for rethinking AI trust and safety, with strong relevance for the Global South. (undp.org)
- Trust and Safety in 2025: Strategies & Top Companies – FIOWE: Insights into emerging industry practices, trends, and strategies for digital platform safety. (foiwe.com)
- Raised by the Internet: Technology-Facilitated Child Sexual Exploitation and Abuse in India – ESYA Centre: Evidence-based research on risks, modalities, and policy gaps related to online child abuse and exploitation in India. (esyacentre.org)

These materials can support continued learning for policymakers, practitioners, and advocates engaged in trust and safety, AI governance, and technology-facilitated harm prevention in the Global South and beyond.

1. <https://indiaai.gov.in/hub/safe-trusted-ai>
2. <https://www.undp.org/digital/ai-trust-and-safety>
3. <https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf>
4. <https://africabrief.substack.com/p/study-faults-indias-legal-system>
5. <https://www.esyacentre.org/perspectives/2025/7/02/raised-by-the-internet-the-reality-of-technology-facilitated-child-sexual-exploitation-and-abuse>
6. <https://www.infosysbpm.com/blogs/trust-safety/trust-and-safety-in-the-era-of-generative-ai.html>
7. <https://www.foiwe.com/the-ultimate-guide-to-trust-and-safety-in-the-digital-age-2025/>
8. <https://nasscom.in/ai/pdf/the-developer's-playbook-for-responsible-ai-in-india.pdf>
9. <https://www.telusdigital.com/insights/trust-and-safety/resource/trust-and-safety-2025>
10. <https://indiaai.gov.in/article/ensuring-safety-trust-and-governance-in-the-age-of-ai>

PHOTO GALLERY



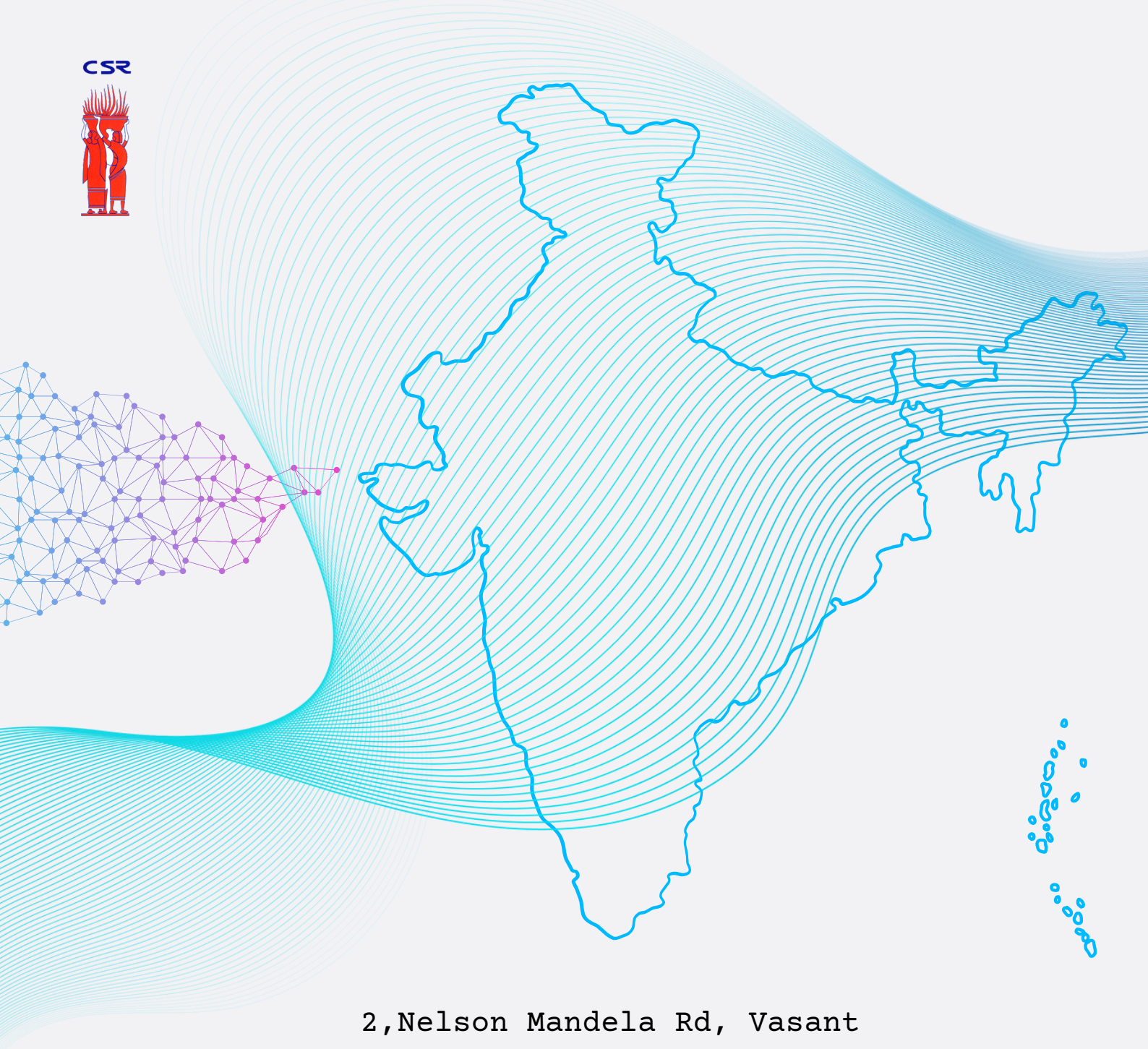




OUR PARTNERS



CSR



2, Nelson Mandela Rd, Vasant
Kunj, New Delhi 110070

www.csrindia.org

+91 011 46131929

info1@csrindia.org



@CSR_India.



@CSR_India.



@Centre For Social
Research (CSR) India



@Centre For Social
Research India



@centresocialresearch