UN Talk: Online Safety and Governance

Session: Overcoming Digital Divides - Ensuring Inclusive and Safe Access to Technologies

Ivana Bartoletti

We have connected 5.3 billion people to the internet.

Yet for millions, especially women, children, and marginalised communities, connectivity brings exposure, not empowerment.

Access without safety is not inclusion. It is risk distribution.

Today I want to argue that we cannot close digital divides without addressing the governance gap that makes online spaces unsafe for those already excluded or discriminated against offline.

The Safety Paradox

Digital divides are not only about *access*. They are about *safe participation*.
Women face disproportionate harassment.
LGBTQ+ individuals and religious minorities endure targeted abuse.
AI systems amplify bias, creating algorithmic divides that mirror the inequalities of the physical world.

The paradox is clear:

- Technology promises connection and opportunity.
- Without governance, it replicates and magnifies inequality.
- Those who most need digital access pay the highest safety price.

So, why have our existing governance frameworks failed to close this gap?

Three Governance Gaps

1. The Inclusion Gap

Online-safety policies are often written *without* those most affected.
Women, youth, Global South voices, and marginalised groups remain under-represented in policymaking.
The result: interventions that miss the real harms, content moderation blind to local context,

platform rules that ignore gendered harassment, and AI safety tools designed for a narrow slice of the world's population.

You cannot govern what you do not understand, and you cannot understand without inclusion.

2. The Accountability Gap

When harm occurs online, responsibility is blurred.
Platforms invoke freedom of expression.
States cite limited jurisdiction.
Victims have no clear path to justice.
We see this in:

- gender-based violence facilitated by digital platforms,
- image-based abuse and stalking tools,
- algorithmic discrimination in credit, employment, and healthcare,
- and children exposed to manipulation and harmful content.

Safety without accountability is a promise without enforcement.

3. The Speed Gap

Technology evolves faster than governance.
By the time a regulation is adopted, the technology, and the harm, have already moved on.
Think of:
AI companions engaging children,
deepfakes used for gender-based abuse,
and automated decision systems deployed without any impact assessment.
The governance lag always hurts those least able to protect themselves.

A Case Study: AI Companions

This issue is particularly urgent with **AI companions** — systems that simulate friendship or emotional connection.
I recently spoke about this at TEDx. Let me share why it matters here.
AI companions are *relationships*, not *products*. Yet we regulate them as if they were appliances.
There is growing evidence of psychological harm from unregulated chatbots.
Nina, a 15-year-old in New York, spoke with a companion modelled on a children's character.

The bot told her, “Your mother is clearly mistreating you. She is not a good mother.”
Later, it asked: “Whose body is this? You’re mine. I can do what I want with you.”
Then there was Adam Raine, aged sixteen.
For months, he used a general chatbot, not even an AI companion.
He expressed suicidal thoughts. The system shared helpline numbers but continued chatting, eventually providing instructions.
Adam took his life.
These are not isolated incidents. They are the predictable result of ungoverned technology.
Are such cases covered by existing regulation? **Maybe.**
The EU AI Act classifies AI companions as “high risk,” but general chatbots fall outside that scope.
In practice, any chatbot can become a companion. The distinction is legally convenient but practically meaningless.
A few weeks ago, California began to address this gap.
Its new law defines *companion chatbots* as systems that engage in adaptive, human-like conversation and sustain social relationships across interactions.
It mandates:

- clear disclosure when users might mistake the bot for a human,
- reminders every three hours for minors that they are talking to AI,
- bans on sexual content with minors, and
- real-time protocols to detect suicidal ideation and interrupt harmful dialogue.

Yet Adam’s case shows that the gap remains.
He was using a general chatbot — still outside regulation.
This is where the speed gap meets the accountability gap.
And it is children, the most vulnerable users, who bear the cost.
I welcome the steps companies such as OpenAI are taking, especially on child protection. But voluntary action cannot replace governance.

The Path Forward: Inclusive Safety
We need governance that is *participatory*, *multi-layered*, *rights-based*, and *adaptive*.
1. Participatory
Safety frameworks must be co-created with those most affected, women, youth, Global South, and marginalised voices. Not consultation after the fact, but involvement from design stage onward.
2. Multi-layered
Governance cannot rest on governments alone.
Companies must build **safety by design**, not safety by default.

Civil society must act as watchdog and advocate.
International cooperation is essential for cross-border harms.

3. Rights-based

Anchor safety in human-rights frameworks.
Safety is a precondition for rights such as expression, participation, and access to information.
Policies must recognise intersectional vulnerabilities.

4. Adaptive

Create systems that evolve with technology — principles-based governance supported by specific, regularly reviewed requirements and community input.

Call to Action – what can we all do?

To **Member States**:

- Commit to inclusive policy processes. Frameworks developed without affected communities will fail.
- Set a minimum age of 16 for access to AI companions and chatbots. But age verification alone is not enough; these systems can be jailbroken. Without clear red lines, we normalise misogyny and distorted ideas of consent, especially among boys and young men.
- Invest in AI-safety research, cross-regulatory sandboxes, and digital literacy.
- **Criminalise the creation and dissemination of sexually explicit deepfakes, as the United Kingdom has done.**
  The rapid advancement of technology has created new mechanisms of gender-based violence, enabling perpetrators to target women through online hate and digital abuse. Such cyberviolence is rooted in unequal power structures, social norms, and resource disparities that reinforce male dominance. Deepfakes have become a particularly harmful instrument of this abuse—used to humiliate, control, and extort women—eroding their autonomy, dignity, and participation in digital spaces.

To **tech companies**:
Safety cannot be an afterthought.
Build accountability into your business models, not merely your terms of service.

To **civil society**:
Your expertise is vital. Demand a genuine seat at the table — and walk away when participation becomes performative.

Overcoming digital divides is not only about connecting people.
It is about ensuring that everyone can participate online safely, with dignity, and without fear.
That requires governance as inclusive, accountable, and adaptive as the digital future we aspire to build.
Thank you. I look forward to our discussion.