Session Outcome Document

Trust Dilemma: AI Across Sectors

Trust Valley

10 July 2025, 17:00 – 17:45

https://www.itu.int/net4/wsis/forum/2025/Agenda/Session/214

Key Issues discussed: Looking Beyond 2025:

- Technology, especially AI, can be both a weapon and a shield. The discussion emphasized the critical need for trustworthy AI that prioritizes security by design, privacy by design, and robust regulation to ensure AI benefits humanity and minimizes harm.
- The proliferation of AI-powered disinformation and manipulation poses a significant threat, particularly for journalists and public perception. The development of "synthetic soldiers" capable of creating vast networks of fake profiles highlights a silent cognitive warfare.
- Cybercriminals increasingly target small and medium-sized enterprises (SMEs) within supply chains to access larger corporations and governments, revealing a critical need for enhanced cybersecurity measures across the entire ecosystem.
- AI development is accelerating at an unprecedented rate, while international policymaking and governance mechanisms struggle to keep pace. This creates a critical juncture where bold decisions and international collaboration are necessary to prevent dramatic events driven by unregulated AI.
- There is a significant gap to bridge in public education regarding the risks and proper use of AI, even among children who are growing up in an AI-pervasive world without adequate guidance on its implications.
- Training of the most widely used generative AI models on predominantly US-centric data and knowledge can influence the information and perspectives presented to users. This raises important considerations regarding cultural representation in AI outputs.

Tangible Outcomes of the session

1. Philippe Stoll Senior Techplomacy Delegate at ICRC shared key efforts to protect civilians from harmful uses of digital tools in conflict zones:
    - Presented the work of ICRC on an international treaty for legally binding international rules to ban autonomous weapons targeting people by 2026, as well as The Digital Dilemmas initiative to immerse users in crisis simulations to expose the real-world consequences of digital tech in warfare.
    - Presented a recently published ICRC AI policy to empower in-house teams to explore AI with safety, ethics, and human impact at its core, ensuring AI adoption is guided by the ICRC's core mandate and Fundamental Principles.
2. Irakli Beridze Head of the Centre for Artificial Intelligence and Robotics at UNICRI presented AIPOL (AI for Policing) initiative:
    - A joint project with Interpol funded by the European Commission called, to design a global toolkit for the responsible use of AI by law enforcement and translate in national standard operating procedures within pilot programs in Brazil, India, Nigeria, Kazakhstan, and Oman.

3. Stéphane Koch, Vice-President of the Board of Directors ImmuniWeb SA demonstrated the dual-use nature of AI technology and presented his work on
    - AI platform to conduct attack surface monitoring and automated penetration testing to enhance information security, digital communication, and online reputation management.
4. Lennig Pedron CEO at Trust Valley emphasized the fact that Cybercriminals increasingly target small and medium-sized enterprises (SMEs) within supply chains to access larger corporations and governments, and announced Trust Valley work in that regard:
    - Trust4SMEs program in collaboration with the State of Vaud dedicated to strengthening SMEs cybersecurity.
    - GovTech B2G platform in collaboration with the World Bank and the Department of Economy at the national level of Switzerland (SECO) to facilitate implementation of new digital technology projects by startups with governments. Ghana is the first country to launch a pilot innovation challenge.

Key Recommendations and Forward-Looking Action Plan for the WSIS+20 Review and Beyond

Agreements/Commitments as an outcome of the session:

- Accelerate AI Governance and Regulation Globally: Prioritize and expedite the development and adoption of robust international governance instruments and regulatory frameworks for AI, acknowledging the rapid pace of technological advancement and ensuring consensus among UN member states. This includes learning from existing models like the EU AI Act.
- Invest in Public Education and Awareness on AI Risks: Implement comprehensive educational programs from an early age, even in school curricula, to equip individuals with the knowledge and discernment necessary to understand and mitigate the risks associated with AI, especially concerning disinformation and data privacy.
- Strengthen Multi-Stakeholder Collaboration for Responsible AI: Foster and expand collaborative initiatives involving governments, international organizations, law enforcement, the private sector (including SMEs), academia, and civil society to jointly develop and implement responsible AI practices, ensuring human rights compliance and addressing emerging threats like cognitive warfare.
- Prioritize Human-Centric AI Development and Deployment: Ensure that AI development and deployment consistently prioritize human safety, well-being, and ethical considerations, particularly in sensitive areas like law enforcement, humanitarian aid, and military applications, with a strong emphasis on explainability, transparency, and accountability.