POST EVENT REPORT | TRUST AND SAFETY FORUM – TECH AGAINST TERRORISM WORKSHOPS, INCYBER EUROPE, LILLE FRANCE

Date: 02 April 2025

1. Introduction

The digital landscape continues to evolve, creating opportunities for threat actors to exploit technology through new tactics. The hybridisation of online harms—where distinct threats like terrorist and violent extremist content (TVEC), child sexual abuse material (CSAM), image-based sexual abuse (IBSA), and disinformation intersect—represents a complex challenge for policymakers, tech platforms, and civil society.

Tech Against Terrorism (TAT), in collaboration with the Trust and Safety (TS) Forum hosted a series of practical workshops at InCyber Forum Europe held in Lille in April 2025. These workshops fostered collaboration between experts, platforms, and NGOs to explore actionable solutions and create lasting impact.

The primary aim of this initiative was to explore the hybridisation of online harms through practical and interactive workshops, facilitating collaboration between key stakeholders to advance innovative, and technology-driven solutions.

Before the workshops TAT and TS Forum developed the following overarching goals:

- Build cross-sector awareness and partnerships to address hybridised online threats.
- Generate actionable insights and foster collaboration between terrorism, CSAM and IBSA prevention and digital safety experts.
- Develop a series of actionable recommendations for addressing hybrid online threats.
- Produce a post-event white paper outlining workshops findings and solutions to counter hybridisation of online harms and drive further research and policymaking.

2. Overview of workshops

We divided the discussion into three 90-minute workshops under Chatham House Rule. Each workshop entailed 20 minutes of expert discussion, 60 minutes of table discussions followed by 10 minutes of a table summary read out. In each workshop, there were 40 participants which were split to 4 tables of 10 respectively. Each table included at a minimum one expert focused on the workshop theme.

Workshop 1 - Understanding and Addressing the Hybridation of Online Threats

Facilitator: Amazon Web Services (AWS)
Topic: Cross-sectoral cooperation frameworks for preventing online threats.

In this workshop experts and participants examined the intersection of various online harms including TVEC, CSAM, gender-based violence (incel movement) and other online threats like disinformation and exploitation of Generative Artificial Intelligence (GenAI). Participants collaborated to explore innovative strategies to counter these hybrid threats in an evolving digital landscape.

Workshop 2 - Enhancing Information-Sharing Through Technology

Facilitator: Cybersecurity Advisors Network (CyAN)
Topic: Generative Artificial Intelligence & Emerging Threats

In this workshop, experts and participants explored cutting-edge tools and technologies for improving cross-platform coordination in addressing hybrid online harms, from TVEC to CSAM, IBSA and gender-based violence. From hashing technologies to signal-sharing tools, participants discussed how to examine technical, operational, and regulatory challenges and proposed actionable solutions.

Workshop 3 - Dignity and Protection: Integrating Survivors of CSAM, IBSA and TVEC

Facilitator: GIFCT - GNET
Topic: Integrating Survivors' perspective, Balancing with Public Interest

In this workshop experts and participants discussed how to incorporate the voices and needs of survivors into efforts to counter CSAM, IBSA and TVEC. They explored survivor-centred approaches to policy and content moderation strategies, balancing the dignity and protection of survivors with legitimate interests, such as public interest.

3. Recommendations

a. Workshop 1 - Understanding and Addressing the Hybridation of Online Threats

i. Key Insights from Panellists' Interventions in Workshop 1

- *GenAI as an Amplifier of Threats*

In a pre-recorded presentation, Tech Against Terrorism addressed how GenAI is lowering the barriers for threat actors to create credible and scalable terrorist propaganda, violent extremist

content, and manipulated material to amplify harm. It stressed the necessity for scalable detection and moderation frameworks.

- *Jurisdictional Complexity*

It was emphasised how difficult building responses to hybrid threats is due to varying national laws and enforcement capacities. It advocated the need for evidence-led, multi-stakeholder frameworks grounded in international human rights law.

- *Proactive Regulatory and Cooperative Strategies*

The group encouraged proactivity through red teaming, open-source tool innovation, robust legal thresholds (e.g., internationally recognised designation lists), and responsible regulatory engagement. The need to redefine terrorism (and violent extreme content), understand the online and offline impact and share TVEC and CSAM data with regulators in real time was also stressed. The deployment of a common online platform to share such insights was presented as a key strategy to break the silo-approach and foster cross-sectoral collaboration.

- *The Role of AI in Adversarial Shift Detection*

It was explored how AI could enhance the detection of content evolution as threat actors adapt and called for continuous iteration and ethical AI safeguards.

- *Disinformation and Identity Manipulation*

A case study was presented on how platform algorithms can foster online radicalisation pathways through exposure to identity-based conspiracies, such as the “Great Replacement” theory. The case highlighted a critical link between algorithmic content delivery systems and the escalation of ideological polarisation, illustrating how design choices in digital platforms can propagate violent extremist narratives online, reinforce belief echo chambers and lead to violent action.

ii. Key Recommendations for Workshop 1

Table 1 – Cross-Sectoral Cooperation

- Establish international regulatory coherence and compliance economies to reduce discrepancies in enforcement.
- Create shared evidence-gathering standards to ensure safe, legal, and actionable interventions.
- Develop multi-stakeholder protocols for rapid response to emerging online harms.

Table 2 – Platform Design & Disinformation

- Encourage transparency in algorithmic recommendations and content amplification systems.
- Train moderation teams to flag and de-escalate identity-based misinformation.
- Engage in user education initiatives to improve digital literacy around conspiracy narratives.

Table 3 – AI and Threat Detection

- Invest in AI systems that can monitor adversarial shifts in threat actor tactics.
- Build ethical guardrails into AI development processes to prevent abuse at scale.
- Promote open collaboration between law enforcement and technologists for innovation sharing.

Table 4 – Regulation & Hybrid Threats

- Harmonise regulatory definitions across CSAM, terrorism, and disinformation to reflect hybrid realities.
- Create shared policy toolkits for smaller platforms to enforce effectively.
- Collaborate with civil society to ensure regulatory measures are rights-respecting.

iii. Conclusion Workshop 1

Workshop 1 highlighted the growing blurring of boundaries between different forms of online threats —ideological, criminal, and exploitative—and reinforced the need for a unified, flexible, and ethical approach to tackling them.
Such an approach will make it possible to develop actionable strategies that uphold human dignity and resilience in the face of evolving online threats. Further collaboration is encouraged to ensure continued momentum in addressing hybrid online harms.

b. Workshop 2 - Enhancing information sharing through technology

i. Key Insights from Panellists' Interventions in Workshop 2

- *GenAI to proactively anticipate upcoming forms of harms*

Working with AI could help to understand how malicious actors use it to disseminate harmful narratives. This could be achieved by using AI to generate different types of problematic content, analysing the process and predicting future production. This could enable different entities to compile a database of harmful content and monitor what is being created by using the database as a hash list.

- *AI, hash databases and the need for more transparency on community guidelines*

It is important for hash databases to include a mix of human moderation and machines to control the potential of infringing content and avoid over moderation, in compliance with community guidelines. This could lead to improved recognition of harmful content by platforms. Currently there is an inconsistency between platforms who use community guidelines which provides a challenge to consistently identify harmful content. There is a need for more concrete community guidelines across platforms.

- *Geopolitical implications*

Moderating political content has a specific challenge as it is important to take the country's political situation into account as well as historical references. Potential solutions discussed were to create country based or European regulations which could be expanded more globally. It is also important that a consensus is found on the bigger issues that are commonly agreed, such as CSAM content, before trying to find responses to other online harms.

ii. Key Recommendations for Workshop 2

Table 1 - Moderation

Addressing common moderation rules on a European scale, beginning with the most consensual issues, such as CSAM. Followed by community guidelines to be adopted by the industry.

Table 2 - Identification of content

Building infrastructure on the identification of content has yet to be done, and the following challenges will have to be addressed:

- Trust in the organisation that manages the infrastructure is essential
- Criteria of the content that will be hosted, including privacy issues
- Content classification depends on companies and laws
- Sharing patterns that help identify abusive or illegal content could include how hashing technology compares to AI (and how the different hashing technologies compare among them).
- Oversight board on moderation policies is a good practice
- Political landscape plays an important role
- Funds to manage this infrastructure

Table 3 - Cross platform collaboration

Not all platforms are equal as currently large platforms receive more attention due to their size, while smaller platforms struggle with resources. It is vital to develop initiatives where all platforms regardless of size can collaborate to ensure better community guidelines for moderation.

Table 4 - Human and Machine Learning

AI technology is increasingly used to assist in the vast amount of content moderation. However, it has its limitations, and it is important that platforms and third-party moderation companies use a mix of human and AI technology to moderate harmful content.

iii. Conclusion Workshop 2

Workshop 2 discussion focused on the need for greater collaboration among platforms to share signals related to different types of online harm and use AI technology to quickly identify and remove such content.

c. Workshop 3 - Dignity and Protection: Integrating Survivors of CSAM, IBSA and TVEC

i. Key Insights from Panellists' Interventions in Workshop 3

- *Harmful content to users is widely spread online*

AI and technologies in general are used to spread gender-based violence and harm on a bigger scale, publicising victims of GBV. Slang and Algospeak efficiently defeat moderation and machine recognition but also create communities of violent actors, coordinating violence under the radar.

- *Many vulnerable groups are targeted to cause harm*

Public figures, especially women, are particularly vulnerable to gender-based violence, including deepfakes depicting them in sexual situations, sexual harassment, bullying and other abuses.

- *Secondary victimisation online happens*

Victims of a terrorist attack may face additional online victimisation as videos and images of the attack with the victim's identity are included in media coverage and social media posts.

ii. Key Recommendations for Workshop 3

Table 1 - Technology and Online Harm

- Technology can be both a tool and a source of harm. For instance, cyber harassment is persistent and inescapable, making it more than a tool—it's a form of continuous online violence.
- The nature of online environments—viral, borderless, and amplified in scale—intensifies the impact of harm.
- "Self-generated CSAM" (e.g., sexting by minors) presents a classification challenge, as traditional legal frameworks are not equipped to address such content.
- Deepfakes, particularly pornographic ones, are created with the explicit intent of online dissemination and further harm.

Table 2 - Terminology and Language

- Language deeply influences perception and policy. Terms like "revenge porn" reflect a problematic, male-centred framing that can exclude or misrepresent victims.
- Misalignment between policy terminology (e.g., from the European Commission) and national legal definitions (e.g., French law) creates barriers to enforcement and victim support.
- "Algospeak" – the use of coded language to circumvent automated content moderation efforts – is becoming more prevalent.

Table 3 – Survivors' Inclusion

- Survivor engagement is essential to ensure that language, policies, and technological solutions are relevant and effective.
- One recurring theme was: "Nothing about us without us" – advocating for survivors' consultation not just in storytelling, but in system and product design.
- There is recognition that survivors should have the opportunity to be part of developing AI and tech solutions that aim to protect them.

Table 4 - Regulatory and Platform Responses

- Regulators such as OFCOM highlighted efforts to include survivor voices in legislation drafting and red-teaming exercises, although only a small percentage of proposed changes were ultimately accepted.
- OFCOM referred to "Online Gender-Based Harm" and emphasised the value of survivor consultations in shaping regulatory efforts.
- ARCOM, in contrast, focuses more on evaluating whether platforms deploy trust-building tools (e.g., trusted flaggers, age verification) rather than tackling specific types of harm.
- There is no legal obligation for platforms to remove all forms of harmful content (e.g., masculinist content), but a "best effort" expectation exists.

iii. Conclusion Workshop 3

In workshop 3 it became apparent that the need for victim voices regardless of the online harms taking place is needed not only to build awareness but for the industry to better understand the crime and its consequences.

4. General Summary

- The hybridisation of online threats—where actors increasingly blend harmful content and techniques, disinformation, and the misuse of AI—demands an equally hybrid and evolving response framework that integrates technological innovation, legal harmonisation across jurisdictions, and survivor-informed policy design.
- It is important to increase initiatives for sharing signals and tools among companies which can increase brand responsibility. It is also important to learn from existing projects instead of reinventing the wheel.
- Several discussions stressed the need for a collaborative, ecosystem-based approach to prevention, including capacity building for law enforcement, NGOs, and tech companies
- Justice design was identified as a key area, noting that victims often lack awareness of their rights and experience re-victimisation through uncoordinated law enforcement responses.

- The inclusion of victims' voices—regardless of the type of online harm—is essential not only to raise awareness but also to help the industry better understand the nature of these crimes and their impact.
- Participants coming from different backgrounds (platforms, regulators, NGOs and solution providers) agreed on the value of organising a follow-up workshop on that would be more oriented on sharing patterns, with a cross-harms approach (CSAM, IBSA, TVEC) and with technology approach (AI and hashing)

Participating Companies

ALMOND
AUTORITÉ DE RÉGULATION DE LA COMMUNICATION AUDIOVISUELLE ET NUMÉRIQUE (ARCOM)
AUTORITEIT ONLINE TERRORISTISCH EN KINDERPORNOGRAFISCH MATERIAAL (ATKM)
AMAZON WEB SERVICES (AWS)
CHECKSTEP
CHRISTCHURCH CALL FOUNDATION
CITCO
COIMISIUN NA MEAN (CNAM)
CYBERPEACE INSTITUTE
CYBERSECURITY ADVISORS NETWORK (CyAN)
GENDARMERIE
GLOBAL FORUM ON EXTREMISM AND TECHNOLOGY (GNET)
GOOGLE
IMAGE ANGEL
M & C SAATCHI WORLD SERVICES
MATHIAS AVOCAT
UK'S OFFICE OF COMMUNICATIONS (OFCOM)
RESOLVER
REVONTULET
SURVIVORS & TECH SOLVING IMAGE-BASED SEXUAL ABUSE (STISA)
TECH AGAINST TERRORISM (TAT)
TIKTOK
TRUST & SAFETY FORUM (TS FORUM)
UNIVERSITÉ BORDEAUX MONTAIGNE
VERIFYMY
VIDENTIFIER
VYANAMS