Moderation and Compliance (1): Challenges Facing Small Platforms

Dimitra Chondrogianni, GRECA
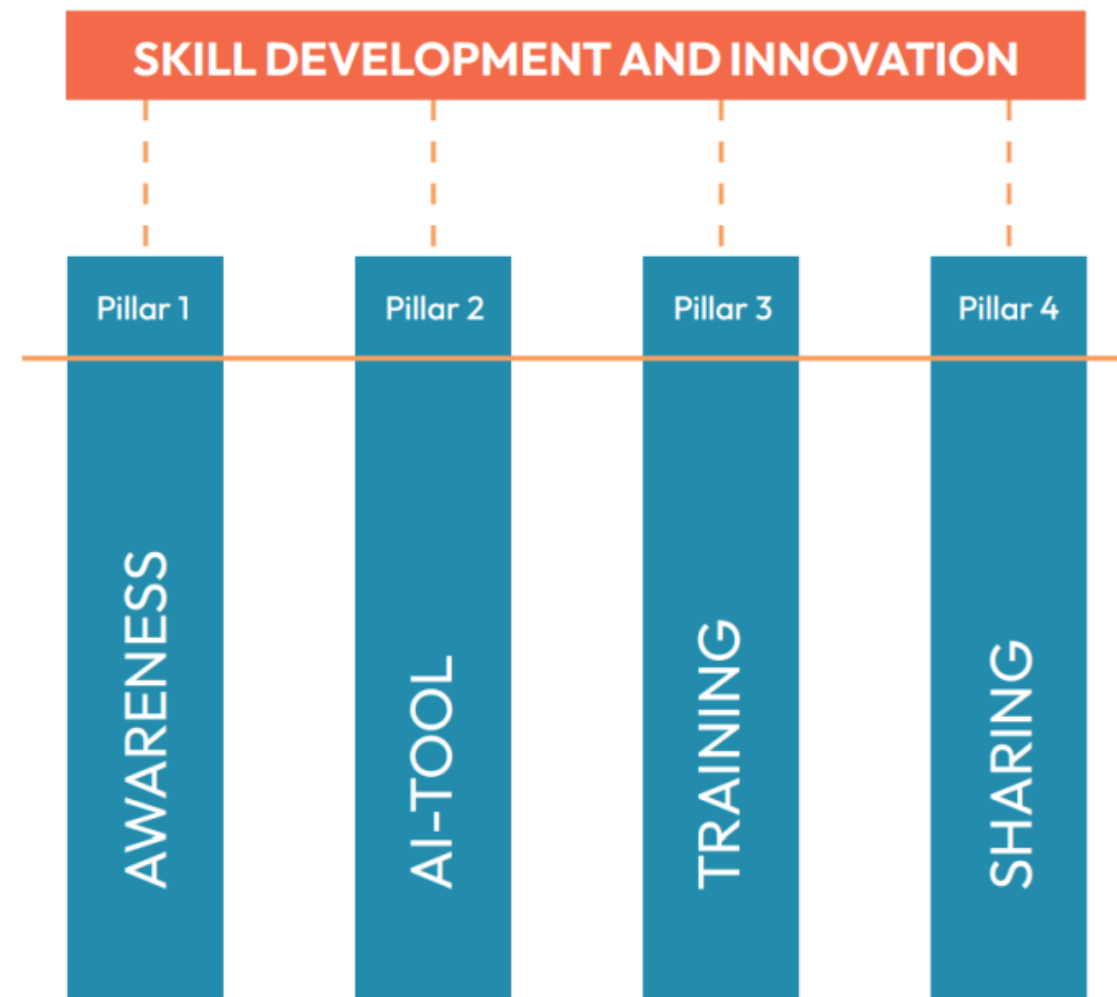
18 October 2024, Brussels

ALLIES

ABOUT ALLIES

One of the most dangerous misuses in the Internet concerning public security are the usage of social media and websites for dissemination of terrorist content.

On that matter the European Union adopted the "Terrorist Content Online" (TCO) Regulation which requires Hosting Service Providers (HSP) to remove any illegal terrorist content within one hour, after having received an official removal order.

Therefore, the aim of ALLIES is to support micro and small HSPs in complying with the requirements of this TCO Regulation through the creation of the proper learning, training, experience sharing mechanisms as well as technical (AI) tools.

SKILL DEVELOPMENT AND INNOVATION

Pillar 1	Pillar 2	Pillar 3	Pillar 4
AWARENESS	AI-TOOL	TRAINING	SHARING

FR!SCO

ALLIES

FACTS

Duration 02/2023 to 02/2025

Programme ISF-2021-AG-TCO

Reference 101080090

CONTACT

🌐 www.alliesproject.com

✉️ office@alliesproject.com

🐦 @ALLIES_EU

in @ALLIES EU-Project

MORE INFORMATION

PROJECT CONSORTIUM

12 Partners | 6 Countries (BG, EL, IT, AT, ES, CY) | Coordinator: LAW and INTERNET FOUNDATION

LAW AND INTERNET FOUNDATION · Information Technologies Institute · novelcore · TRANSCRIME · UNIVERSITÀ CATTOLICA del Sacro Cuore · BEST · SYNYO

mossos d'esquadra · Center for Social Innovation · catalink — The catalyst for innovation · GRECA Greek e-Commerce Association · webarts. · upf. Universitat Pompeu Fabra Barcelona

FRISCO

ALLIES

PROJECT OBJECTIVES

AWARENESS

among micro and small HSPs for the TCO Regulation and its requirements

AI TOOL

developed to support HSPs by implementing the TCO Regulation

TRAINING

for HSPs concerning TCO Regulation content and the usage of the created AI tool
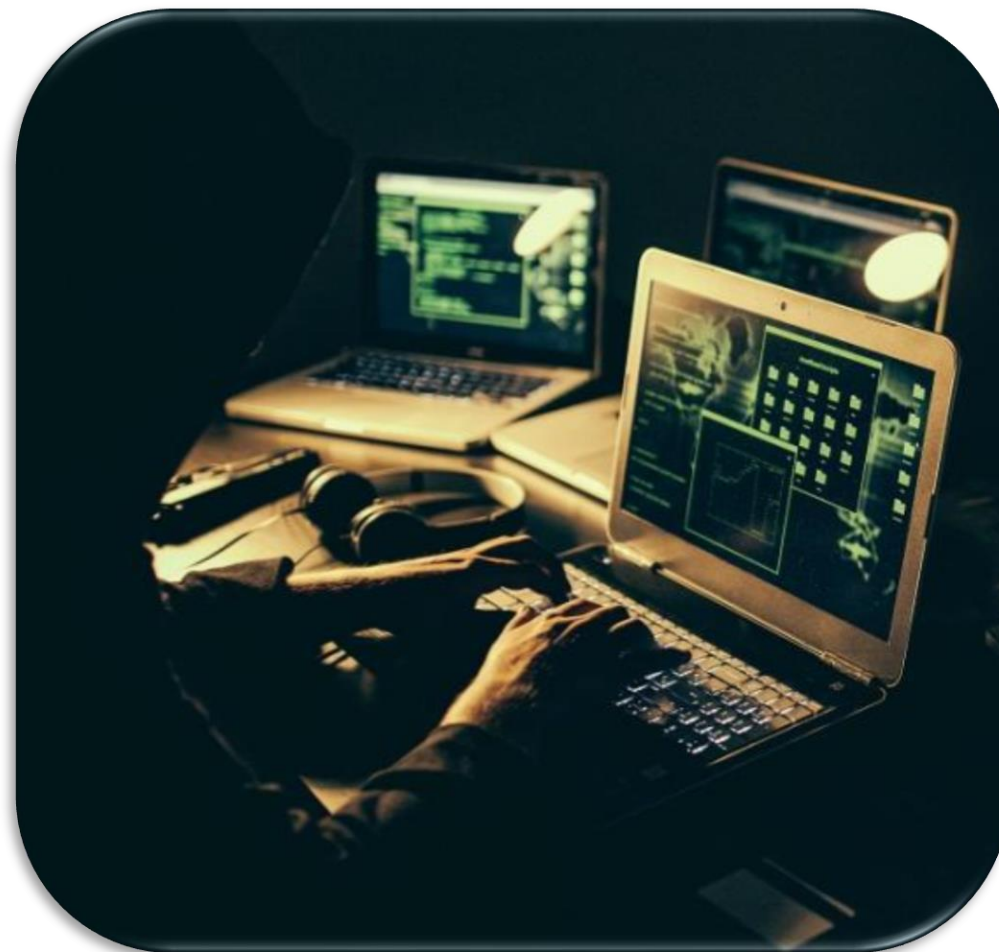
SAFE ONLINE ENVIRONMENT

for experience sharing and reporting among HSPs

FRISCO
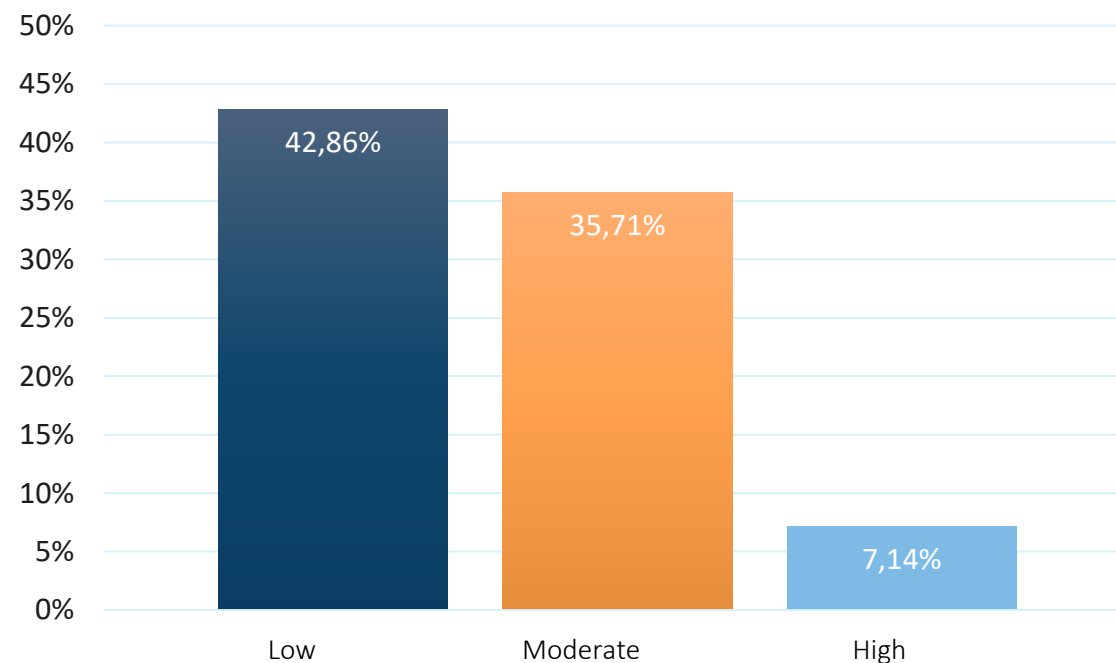
Main challenges faced by small platforms

- **Limited resources**: Small and micro Hosting Service Providers (HSPs) often lack dedicated resources for content moderation.

- **High vulnerability to exploitation**: Smaller HSPs are seen as "softer" targets.

- **Compliance with complex regulations**: Most of the small and micro HSPs lacked awareness of the TCO Regulation.

Main challenges faced by small platforms

➢ **Difficulty implementing regulations:** Due to limited expertise and the cost of implementing compliance measures

➢ **Lack of moderation tools:** A large proportion of small HSPs have no established procedures for dealing with flagged terrorist content. This highlights the need for accessible, cost-effective moderation tools for small platforms.
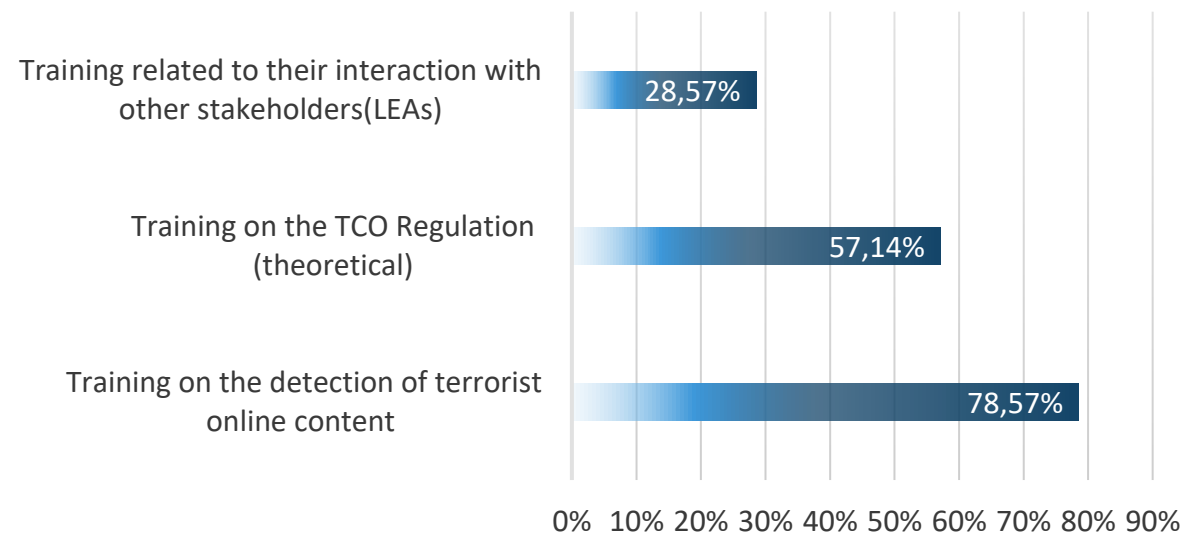
HSPs' expertise in applying the TCO Regulation

Bar chart:
- Low: 42,86%
- Moderate: 35,71%
- High: 7,14%

Main challenges faced by small platforms

➢ **Manual processes:** Being time-consuming, less efficient, and prone to human error

➢ **Training gaps:** Need for training, especially around the detection of terrorist content (78.57%) and understanding the regulatory landscape (57.14%). Developing **user-friendly training materials** is crucial for improving their compliance and moderation efforts.

HSPS' NEED FOR TRAINING

Training related to their interaction with other stakeholders(LEAs) — 28,57%

Training on the TCO Regulation (theoretical) — 57,14%

Training on the detection of terrorist online content — 78,57%

0% 10% 20% 30% 40% 50% 60% 70% 80% 90%

Main challenges faced by small platforms

> **Tool familiarization:** Training to familiarize the platform operators with AI-driven

> **Unified reporting systems:** Standardized risk assessment modules and unified reporting forms for terrorist-related content

> **Continuous threat monitoring:** Threat landscape monitoring and regular updates due to rapidly evolving threats

ALLIES AI tools

ALLICS

TEXT ANALYSIS
Processing of multilingual textual content hosted online by HSPs

SPEECH RECOGNITION
Multilingual automatic speech recognition of the spoken content hosted online

VISUAL UNDERSTANDING
Object and human behavior recognition on images/ videos

MULTIMODAL DATA HASHING
Information retrieval of duplicate or similar content based on hash representations

MULTIMODAL CLASSIFICATION
Categorizing multimodal content based on its relevance to TCO

KNOWLEDGE GRAPH
Decision support based on Knowledge Graphs, creating semantic relations

EXPLANATION ENGINE
Diverse explainability & reporting based on AI

RISK ASSESSMENT
Identifying and assessing the level of risk of content HSPs in terms of TCO

FEDERATED LEARNING
Decentralised machine learning (ML) approach ensuring privacy & security

FR!SC⊖

ALLIES

- Compliance to **TCO Regulation** with limited resources
  - Response to **removal orders** within the golden hour

- **Proactive removal** of TCO
  - HSP content cross-checked against the **TCO hash-based repository**
    - **Validated TCO content** automatically removed
  - Automated content processing for the **detection** and **identification** of TCO
  - **Human-comprehensible explanations** for **content moderation**

- Voluntary **proactive reporting** of TCO

- **On-premises** deployment or on the **cloud**

Federated AI model **(RE-)TRAINING** without sharing raw data

FRISCO

Thank you!