ALLIES

FR!SCO

Fighting Terrorist Content Online

tech against terrorism europe

Terrorist Content Online (TCO): how to prevent it?

How terrorist organisations spread TCO and what can be done to stop it

11.10.2023

10:00 – 11:30 AM CEST

Host: SYNYO GmbH – ALLIES LEAD

Housekeeping Rules

This session will be entirely recorded and published on the ALLIES and/or FRISCO, TATE website.

Please mute your microphone while not speaking due to background noise.

Feel free to post your questions in the **chat**.

If you would like to speak, raise your hand and wait for the moderator to give you the floor.

Meet the
TCO Cluster!

Working together to raise awareness
of the Terrorist Content Online Regulation

ALLIES

FRISCO
Fighting Terrorist Content Online

tech
against
terrorism
europe

Sibling projects under the EU call ISF-2021-AG-TCO

#TCOCluster

TCO Cluster

The **TCO Cluster** aims to inform and support small and micro **hosting service providers (HSPs)** about the new regulation on terrorist content online and their new obligations. **Tools** will be created to report and remove the content, while respecting human rights and fundamental freedoms.

Follow us & stay updated

ALLIES	in @ALLIES EU-Project @ALLIES_EU	alliesproject.com
FR!SCO Fighting Terrorist Content Online	in @FRISCO EU project @FRISCOproject	friscoproject.eu
tech against terrorism europe	in @Tech Against Terrorism @techvsterrorism	tate.techagainstterrorism.org

ALLIES

ALLIES general objectives

- **Increase** awareness among micro and small HSPs for the TCO Regulation and its requirements
- **Develop** AI-based tools for the effective detection and removal of TCO, increasing the capacity of HSPs to comply with the TCO Regulation
- **Train** HSPs on TCO Regulation content and the use of the developed AI-based tools
- **Create** a safe online environment for experience sharing and reporting among HSPs

FRISCO

Fighting Terrorist Content Online

FRISCO general objectives

- **Inform** HSPs and increase their awareness of the Terrorist Content Online Regulation as well as their new obligations.
- **Develop** and validate tools, frameworks and mechanisms to support HSPs in the implementation of the TCO Regulation.
- **Share** experience, best practices and tools to support the implementation of the Regulation.

tech against terrorism europe

TATE general objectives

- **Enable** and equip smaller hosting service providers (HSPs) to disrupt and tackle terrorist content online as per their obligations towards the EU’s TCO.
- **Develop** resilience amongst HSPs through mentorship capacity building programmes and bespoke advisory.
- **Expand** tools already developed by Tech Against Terrorism including the Knowledge Sharing Platform and the Terrorist Content Analytics Platform: the world’s largest alerting database of verified terrorist content.
- **Drive** awareness of the EU's Terrorist Content Online Regulation by convening TATE's consortium of experts and networks from academia and civil society.

Terrorist Content Online (TCO)-Regulation

ALLIES

FRISCO

tech against terrorism europe

LAW AND INTERNET FOUNDATION

RESEARCH CENTER FOR LAW AND INFORMATION TECHNOLOGIES

Martina Bogdanova

Legal Researcher at LIF

ALLIES project

Terrorist Content Online (TCO)-Regulation

LIF: Law and Internet Foundation

- **Main objective**: Reduce the impact of and vulnerability to terrorist content online
- Applies to **HSPs offering services in the European Union**, irrespective of their place of main establishment
- Each Member State has designated **competent authority** to fulfil obligations under the TCO Regulation
- **Exemption** for materials disseminated for educational, journalistic, artistic or research purposes applies

Terrorist Content Online (TCO)-Regulation

LIF: Law and Internet Foundation

Article 2, par. 1 TCO Regulation:

'**hosting service provider**' means a provider of services as defined in point (b) of Article 1 of Directive (EU) 2015/1535 of the European Parliament and of the Council, consisting of the storage of information provided by and at the request of a content provider;

> ***Article 1, p. b) Directive (EU) 2015/1535***:
>
> 'service' means any Information Society service, that is to say, any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services.
>
> For the purposes of this definition:
>
> (i) 'at a distance' means that the service is provided without the parties being simultaneously present;
>
> (ii) 'by electronic means' means that the service is sent initially and received at its destination by means of electronic equipment for the processing (including digital compression) and storage of data, and entirely transmitted, conveyed and received by wire, by radio, by optical means or by other electromagnetic means;
>
> (iii) 'at the individual request of a recipient of services' means that the service is provided through the transmission of data on individual request.

Terrorist Content Online (TCO)-Regulation

LIF: Law and Internet Foundation

***Article 2, par. 7 TCO Regulation**:*

„**terrorist content**“ means one or more of the following types of material, namely material that:

a) incites the commission of one of the offences referred to in points (a) to (i) of Article 3(1) of Directive (EU) 2017/541, where such material, directly or indirectly, such as by the glorification of terrorist acts, advocates the commission of terrorist offences, thereby causing a danger that one or more such offences may be committed;

b) solicits a person or a group of persons to commit or contribute to the commission of one of the offences referred to in points (a) to (i) of Article 3(1) of Directive (EU) 2017/541;

c) solicits a person or a group of persons to participate in the activities of a terrorist group, within the meaning of point (b) of Article 4 of Directive (EU) 2017/541;

d) provides instruction on the making or use of explosives, firearms or other weapons or noxious or hazardous substances, or on other specific methods or techniques for the purpose of committing or contributing to the commission of one of the terrorist offences referred to in points (a) to (i) of Article 3(1) of Directive (EU) 2017/541;

e) constitutes a threat to commit one of the offences referred to in points (a) to (i) of Article 3(1) of Directive (EU) 2017/541;

Terrorist Content Online (TCO)-Regulation

LIF: Law and Internet Foundation

- **Standardised templates** for removal orders *(Annex I)*
- Removal or disabling access to terrorist content **within one hour of receipt** *(confirmation with Annex II)*
- **Exemptions** for non-compliance on grounds of:
  - ➢ *force majeure* or *de facto* impossibility
  - ➢ manifest errors or lack of sufficient information for execution *(communicated with Annex III)*
- Applicable procedures and deadlines sent **at least 12 hours before issuing** the first removal order.
- Terms & Conditions shall include information on the misuse of the HSPs' services for dissemination of terrorist content if they have been exposed to it

LEAs Perspective

Strengths and difficulties in applying the TCO regulation and in detecting and removing terrorist content online from the LEAs

Agent Baez

Member UCIRAX at Mossos Esquadra

ALLIES project

LEAs Perspective

INT: mossos d’esquadra

MOSSOS D’ESQUADRA - CGINF - CENTRAL AREA INFORMATION TECHNOLOGIES - UCIRAX

UCIRAX

Central Investigation Unit for Online Radicalisms

- Scanning the internet to detect and respond to terrorist and violent extremist content online
- Provide support on CT internet-based investigations.
- Apply the EU CRISI Protocol (EUCP) on a collective response to the viral spread of terrorist and violent extremist content online.
- Participate in tasks (flagging) of online illegal content in cooperation with a network of national counterparts.
- Providing prompt and effective support in close cooperation with the organizations providing the websites to remove such content.

Technologies:

- OSINT/SOCMINT.
- SNA.
- Automatic video and audio image processing in the counter-terrorism context.
- Computer Vision.
- Entity Extraction an NPL.
- AI.

LEAs Perspective

INT: mossos d’esquadra

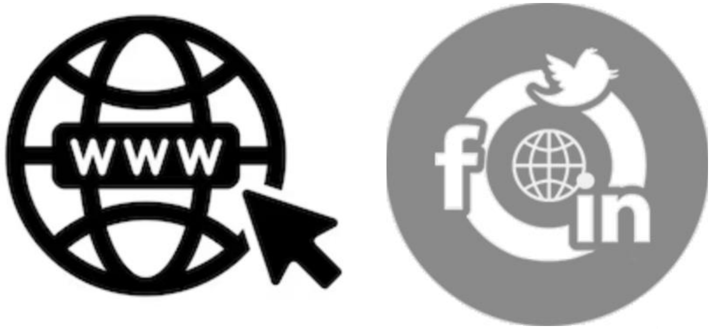
Online Ecosystem

LEAs Perspective

INT: mossos d’esquadra

LEAs Perspective

INT: mossos d’esquadra



LEAs Perspective

INT: mossos d’esquadra

LEAs Perspective

INT: mossos d’esquadra

LEAs Perspective

INT: mossos d’esquadra

LEAs Perspective

INT: mossos d’esquadra

LEAs Perspective

INT: mossos d'esquadra

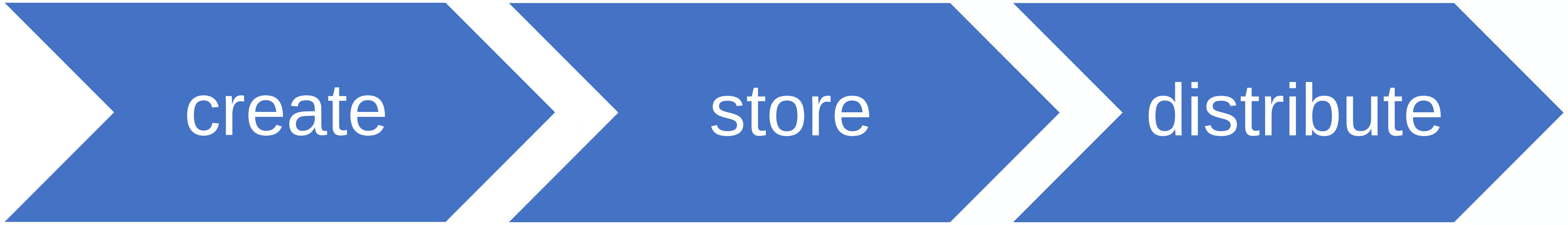
LEAs Perspective

INT: mossos d'esquadra

Online Ecosystem

- Each services can be classified in 2 groups based on how they are being used.

  - VESSELS
  - DISSEMINATORS



LEAs Perspective

INT: mossos d’esquadra

Online Ecosystem

- VESSELS:

Online services used by terrorist actors to host the generated content or propaganda with the aim of distribute it later on using disseminators or other means.

LEAs Perspective

INT: mossos d’esquadra

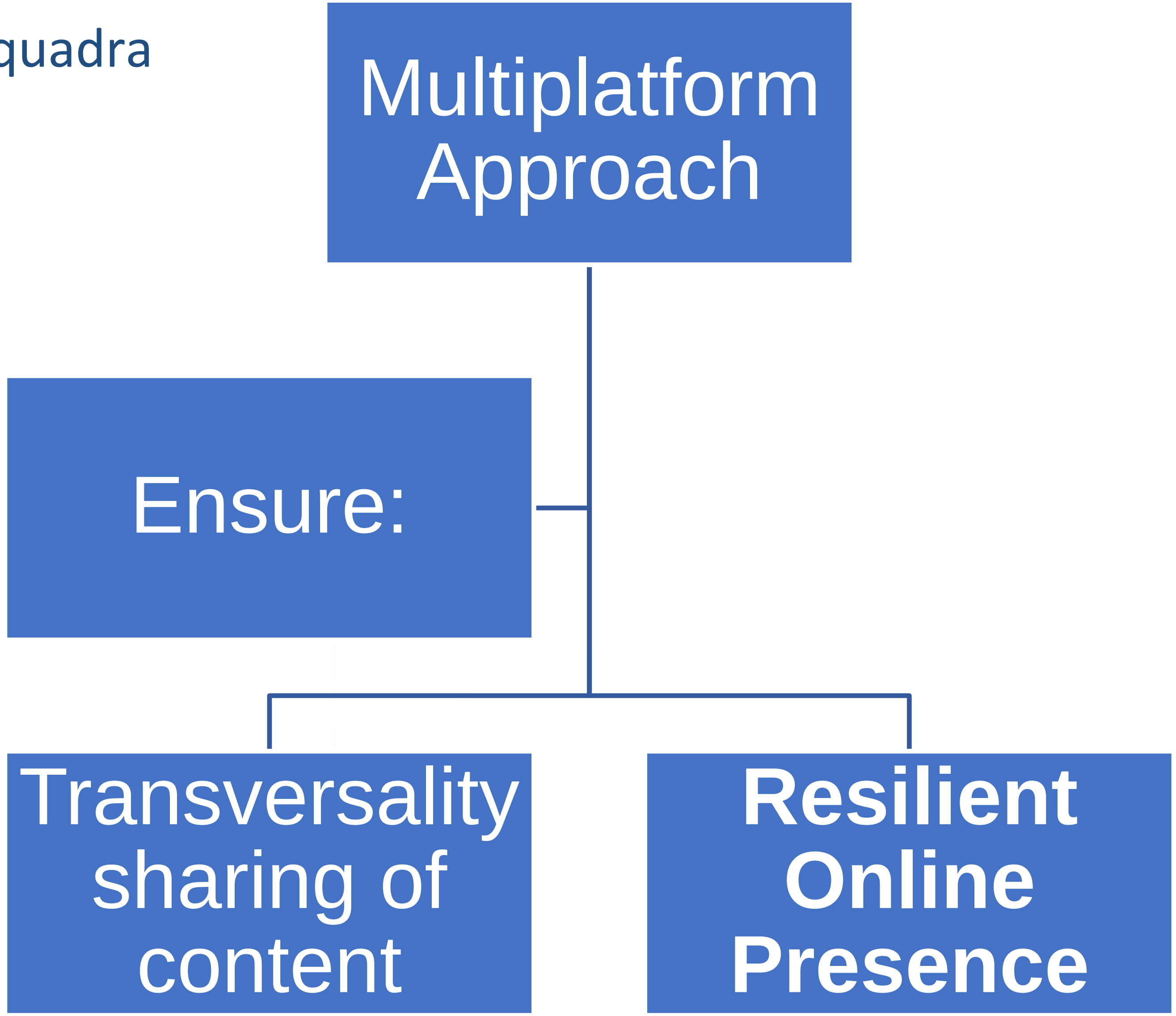
Online Ecosystem

- DISSEMINATORS:

Online services used by terrorist actors to spread generated content or propaganda with the aim of reaching the largest audience possible based on a specific selected objective.

LEAs Perspective

INT: mossos d’esquadra



Multiplatform Approach

Ensure:

Transversality sharing of content

Resilient Online Presence

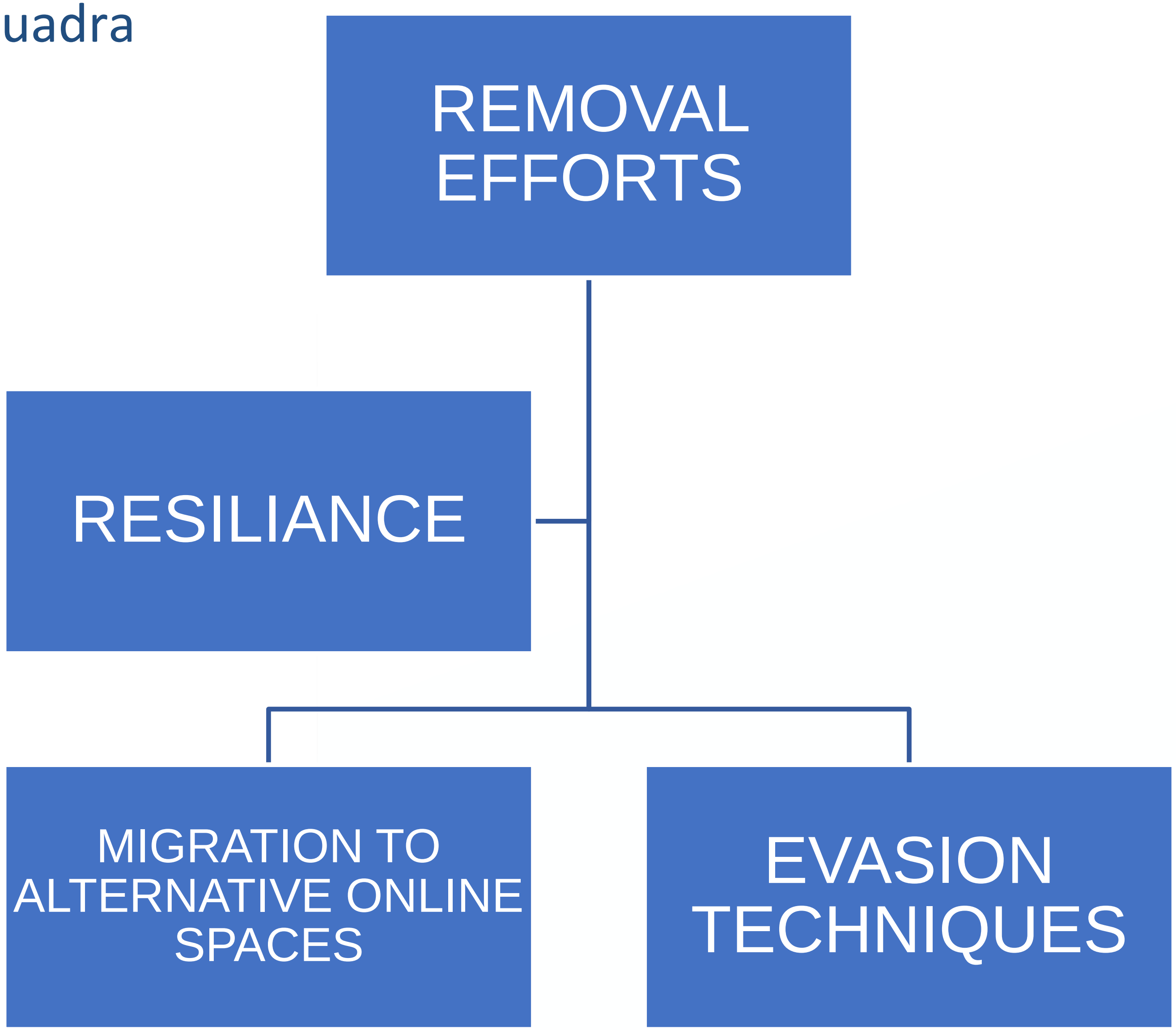
LEAs Perspective

INT: mossos d’esquadra

RESILIENT

LEAs Perspective

INT: mossos d’esquadra



REMOVAL EFFORTS

RESILIANCE

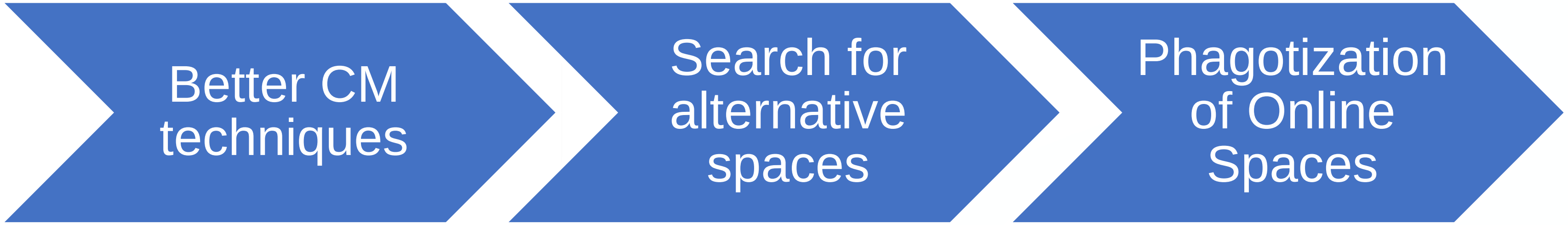
MIGRATION TO ALTERNATIVE ONLINE SPACES

EVASION TECHNIQUES

LEAs Perspective

INT: mossos d’esquadra

MIGRATION TO ALTERNATIVE ONLINE SPACES:



Better CM techniques

Search for alternative spaces

Phagotization of Online Spaces

LEAs Perspective

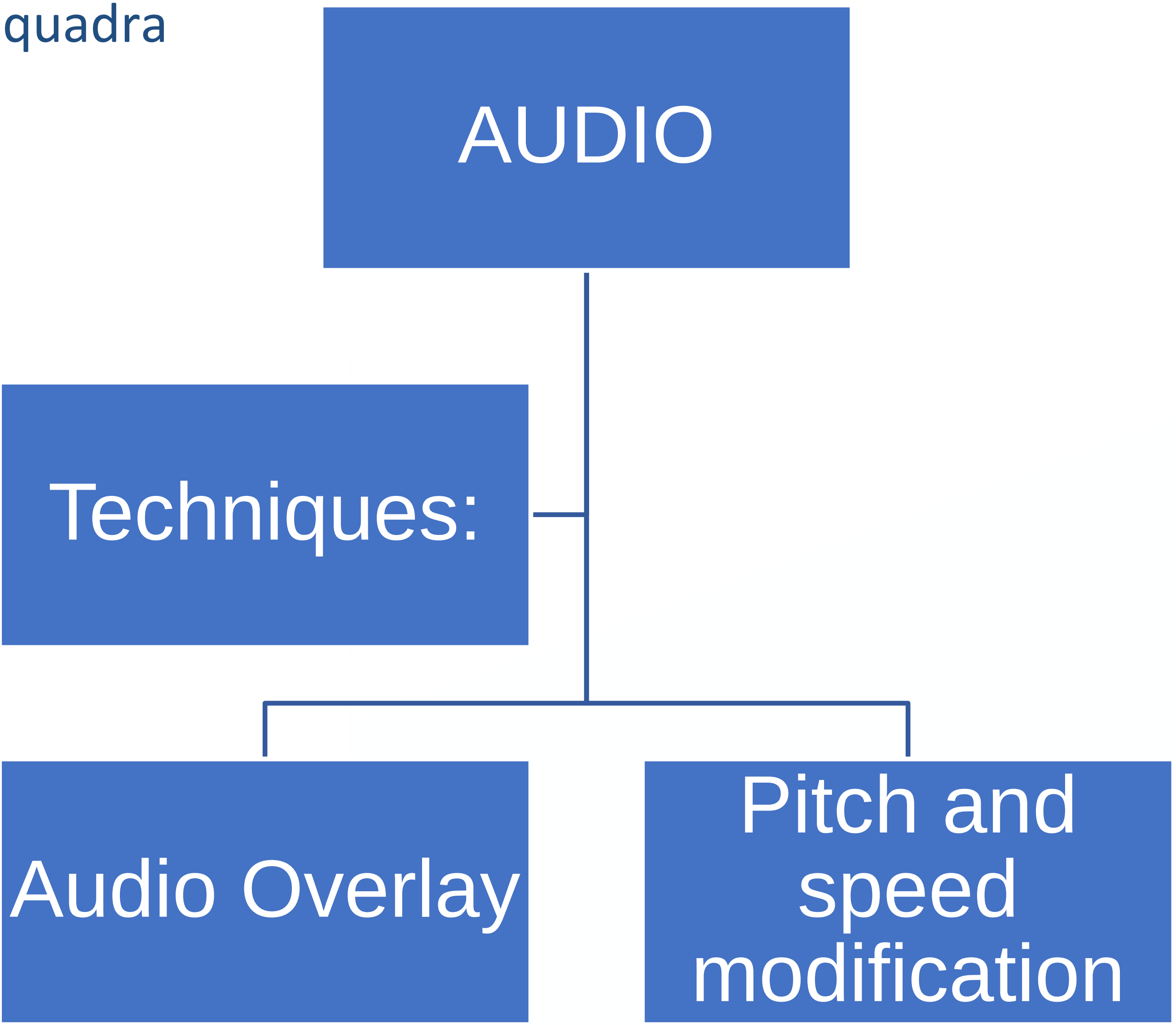
INT: mossos d’esquadra

EVASION TECHNIQUES:

Terrorist actors use a variety of methods to evade content removal. We can classify these techniques according to the type of content: AUDIO, VIDEO (IMAGE) OR TEXT.

LEAs Perspective

INT: mossos d’esquadra



AUDIO

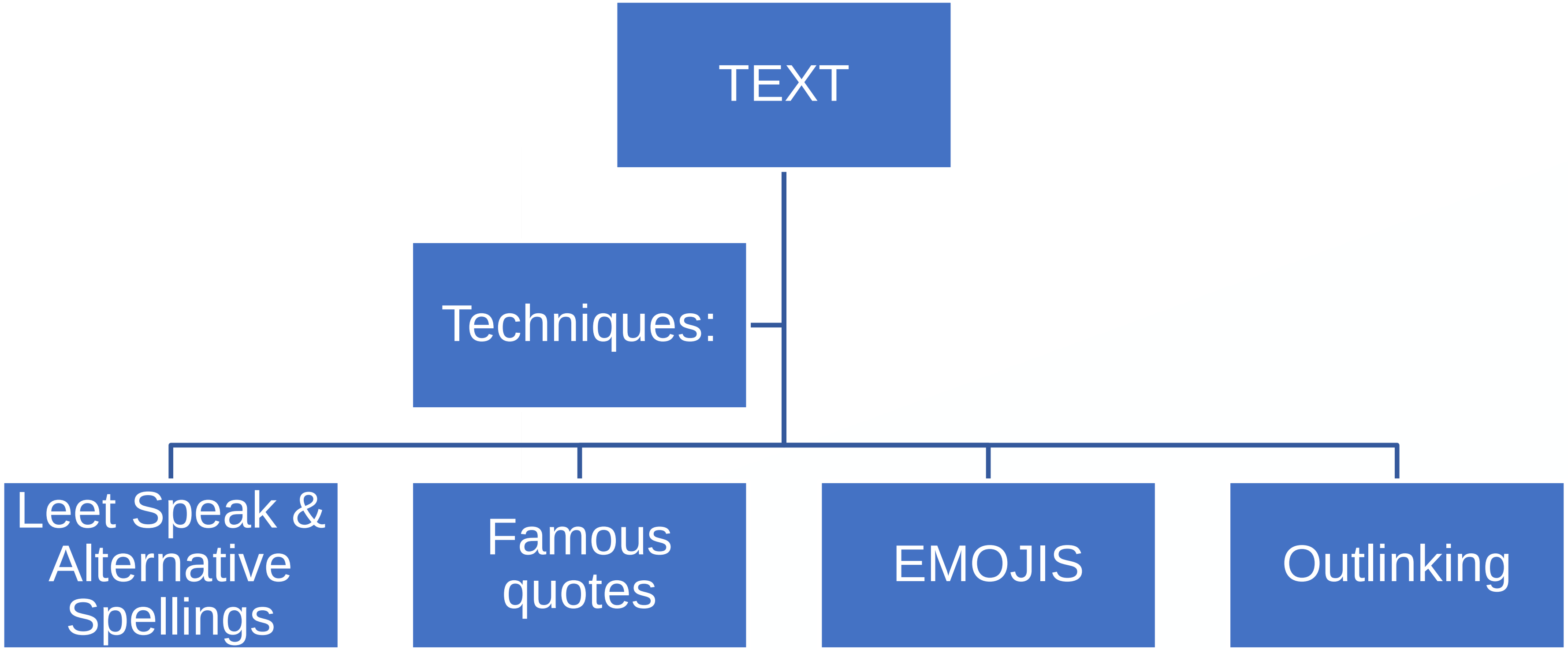
Techniques:

Audio Overlay

Pitch and speed modification

LEAs Perspective

INT: mossos d’esquadra


TEXT

Techniques:

Leet Speak & Alternative Spellings

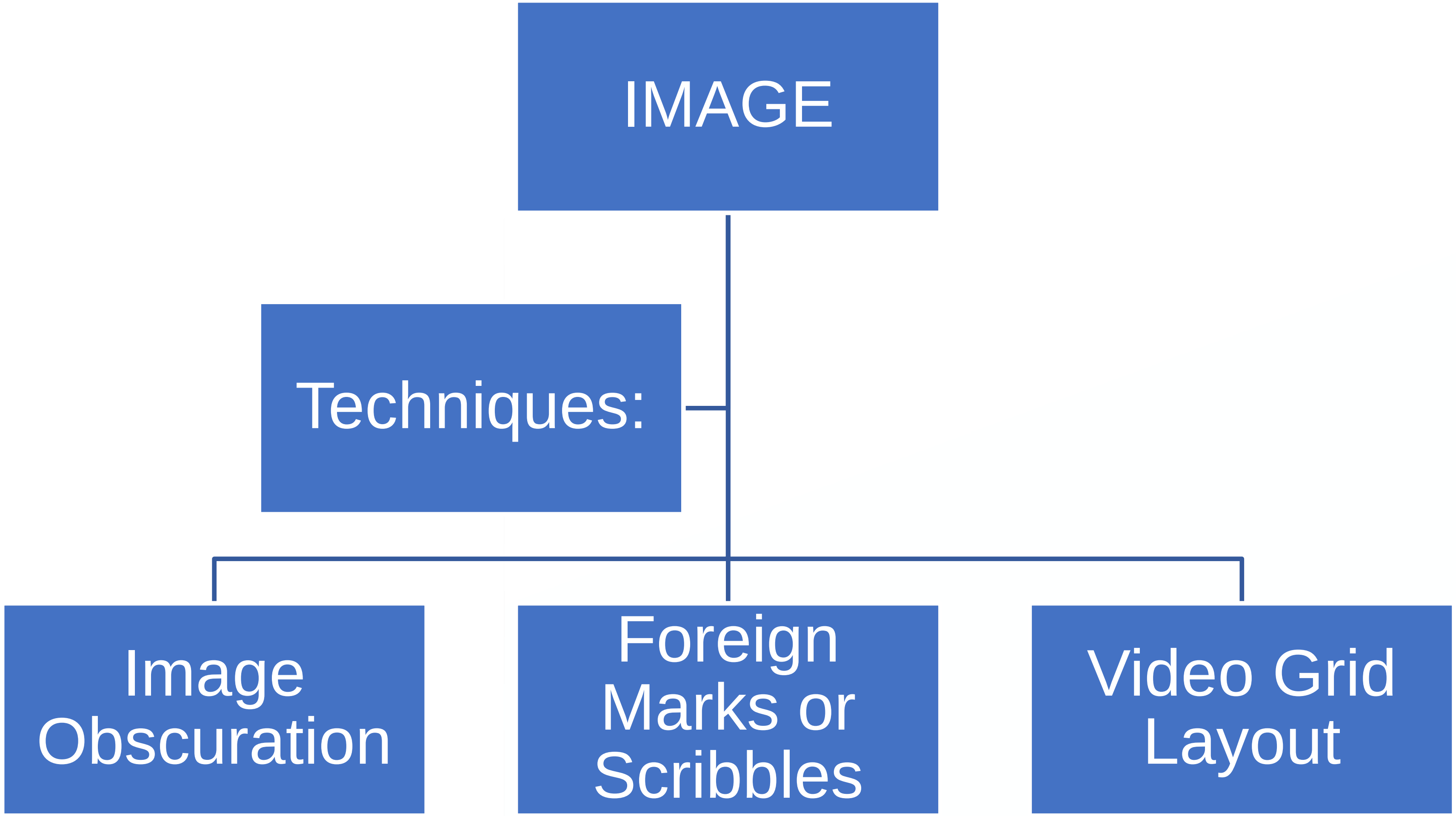
Famous quotes

EMOJIS

Outlinking

LEAs Perspective

INT: mossos d'esquadra



LEAs Perspective

INT: mossos d’esquadra

GRÀCIES – GRACIAS – THANK YOU

Generalitat de Catalunya

mossos d’esquadra

HSP Perspective

Strengths and difficulties in applying the TCO regulation and in detecting and removing terrorist content online from the Host Service Providers

Alexandre Euverte
Data Protection Officer

Margaux Liquard
Head of Trust and Safety

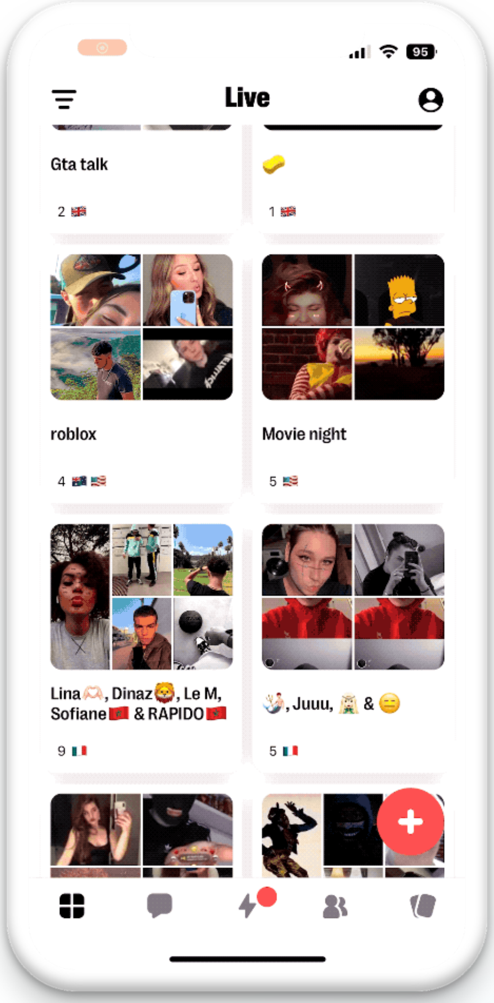
FRISCO project

What is Yubo?

Yubo, the live social discovery platform for GenZ

ALLIES FRISCO tech against terrorism europe

- Yubo is a **French social discovery app**, founded by 3 French engineers in 2015.
- Yubo's core feature is **live video**, allowing users to communicate in small groups.
- **Online safety** has always been a top priority for Yubo, so that users can make meaningful connections and feel confident.



Yubo x The Legal Approach

What is a “terrorist content” legally speaking?

TCO Regulation refers to the definition of **“*terrorist offences*”** in EU Directive 2017/541 on combating terrorism.

→ Very broad definition which may include “*attacks upon a person’s life*” for the purpose of “*seriously intimidating a population*” or “*destabilising the fundamental political, constitutional, economic, or social structures of a country*”.

What content?

“***Terrorist Content Online***” (TCO) includes material that:

- Incites or advocates terrorist offences, such as by glorification of terrorist acts;
- Solicits someone to commit terrorist offences;
- Provides instruction on how to use weapons to conduct attacks;
- make threats to commit terrorist offences.

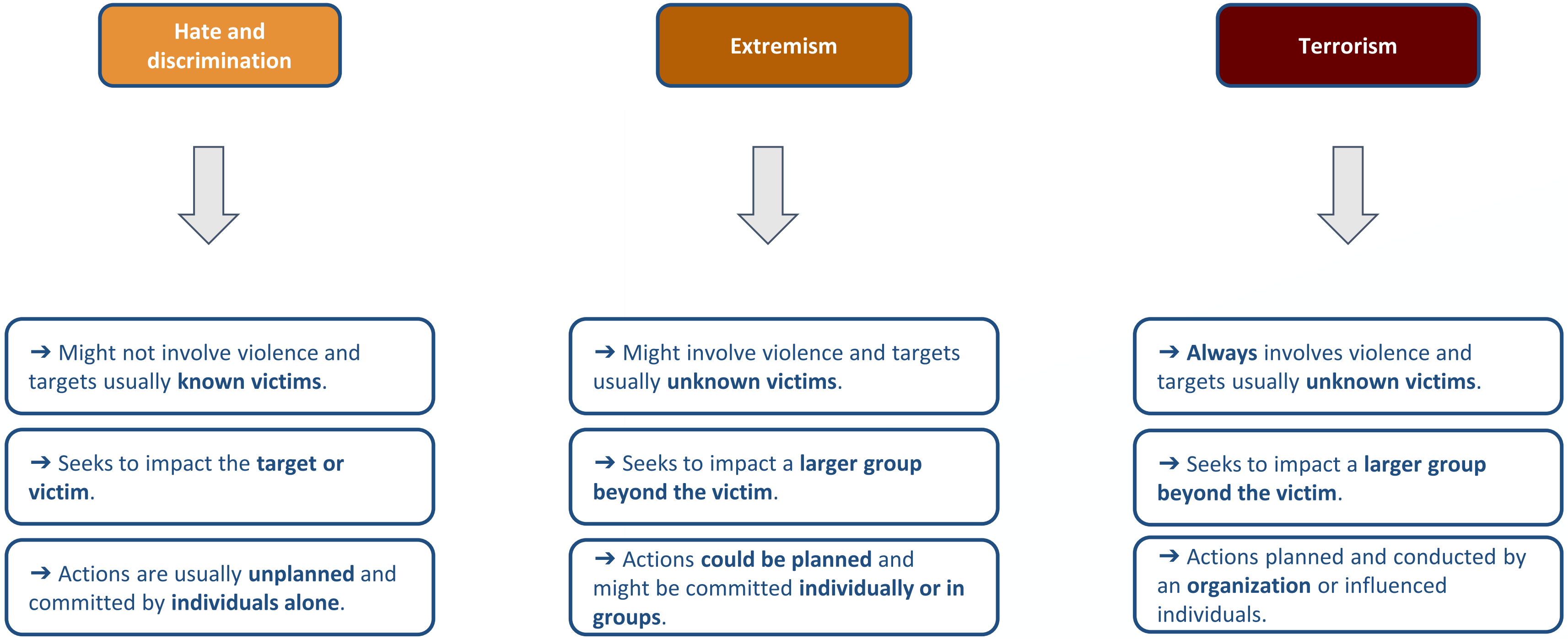
What material?

→ Text, images, sound recordings, videos, live transmission of terrorist offences, etc.

Exceptions: Material disseminated for educational journalistic, artistic or research purposes or awareness-raising purposes.

Fighting extremism, hate and violence

Policy and concepts explanation



Hate and discrimination

➔ Might not involve violence and targets usually **known victims**.

➔ Seeks to impact the **target or victim**.

➔ Actions are usually **unplanned** and committed by **individuals alone**.

Extremism

➔ Might involve violence and targets usually **unknown victims**.

➔ Seeks to impact a **larger group beyond the victim**.

➔ Actions **could be planned** and might be committed **individually or in groups**.

Terrorism

➔ **Always** involves violence and targets usually **unknown victims**.

➔ Seeks to impact a **larger group beyond the victim**.

➔ Actions planned and conducted by an **organization** or influenced individuals.

What are platforms' main obligations?

The TCO Regulation does not call into question the principle of ***‘platform neutrality’***:

No obligation to monitor the content they host.

No obligation to actively investigate illegal activities on their services.

No obligation to use automated tools.

Platforms “*exposed to terrorist content*” may be required to take “***specific measures***” to stop the dissemination of such content. But they can choose which measures to take.

What are platforms' main obligations? (1/2)

#1 Single point of contact (PoC) and legal representative

→ PoC will receive removal orders by electronic means.

#3 Proactive reporting

→ Promptly inform LEA in case of terrorist content involving an *"imminent threat to life"*

#2 Remove / Disable Access

→ Remove or disable access to the terrorist content within **1 hour** of getting a removal order.

#4 Automated tools

→ Implement human oversight and verification procedures for automated detection tools

What are platforms' main obligations? (2/2)

#5 Data preservation

→ Preserve removed terrorist content and related data for **6 months**.

#6 Complaint mechanism

→ Establish processes to challenge a decision to remove or block a terrorist content.

#7 User information

→ Inform users of moderation decision to remove or block content, with the reason for such decision.

#8 Transparency

→ Publish **annual transparency reports** with detailed data on measures taken to address terrorist content and statistics.

Procedures & Penalties

→ HSPs who received a removal order from LEAs have a right to an **effective remedy** to challenge the order.

→ Failure to comply with the TCO obligations can result in penalties up to **4% of the platform's global turnover**.

Yubo x Safety by design

Developing a proactive detection

When it comes to Yubo technology, we detect signals of violence and hate that may be related to extremist /terrorist activity through user content and live streaming.

PHOTOS AND VIDEOS

We detect the presence of a weapon, violence, gory content (such as blood), and hate symbols

TEXT AND STREAMING ACTIVITY

Glorification of hateful or dangerous ideologies, hateful or discriminatory language, and violent threats

The use of proactive **technology** is transparent via multi-layer information:

- Privacy Policy ;
- Contextual information in the app.

Our operational strategy

1) Developing expertises

- We have developed a team of experts focused on high risk and emergency situations, they're trained to identify extremists content and address them accordingly

2) Report priorisation

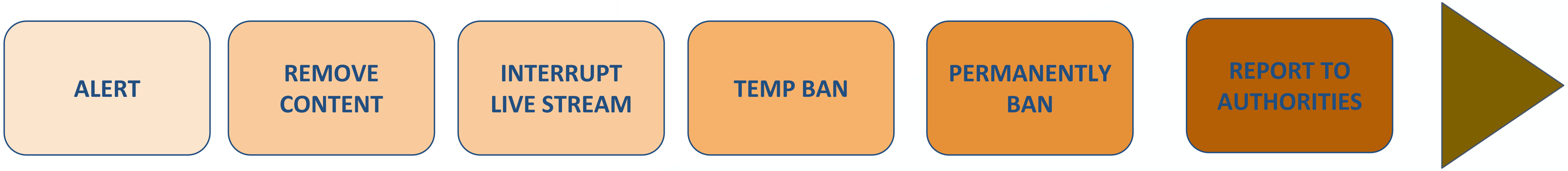
- We prioritise high risk and emergency situations in all of our moderation entry points (Live streaming, profile reported, proactive detection) and set specific t/a time to take action according to the risk reported.

3) Feedback loop from user reports

- We constantly improve our proactive detection by learning from our user report system

Enforcing our policy

- All content related to **hate and extremism** will be used as a signal for our team of specialists to investigate the user account and take appropriate action.
- Depending on the severity, repetition, and urgency of the case, enforcement actions may vary.



ALERT

REMOVE CONTENT

INTERRUPT LIVE STREAM

TEMP BAN

PERMANENTLY BAN

REPORT TO AUTHORITIES

Contextual factors and challenges

When it comes to enforcing our extremism and hate policy on content, contextual factors will determine the level of enforcement actions taken, such as:

- Access to firearms or weapon
- A threat of violence
- An affiliation to a known organisation
- A positive statement towards an extremist ideology

Challenges

- **Operational strain** Investigations can be time consuming which affect resources and costs
- **Lack of evidence** It can be difficult to qualify/map a behaviour from limited user content
- **Legal and privacy issues** Balancing neutrality and privacy with online safety
- **Operational training** Identifying terrorist organisations and figures in a multicultural environment

Collaborating with experts

GIFCT
Global Internet Forum
to Counter Terrorism

tech
against
terrorism

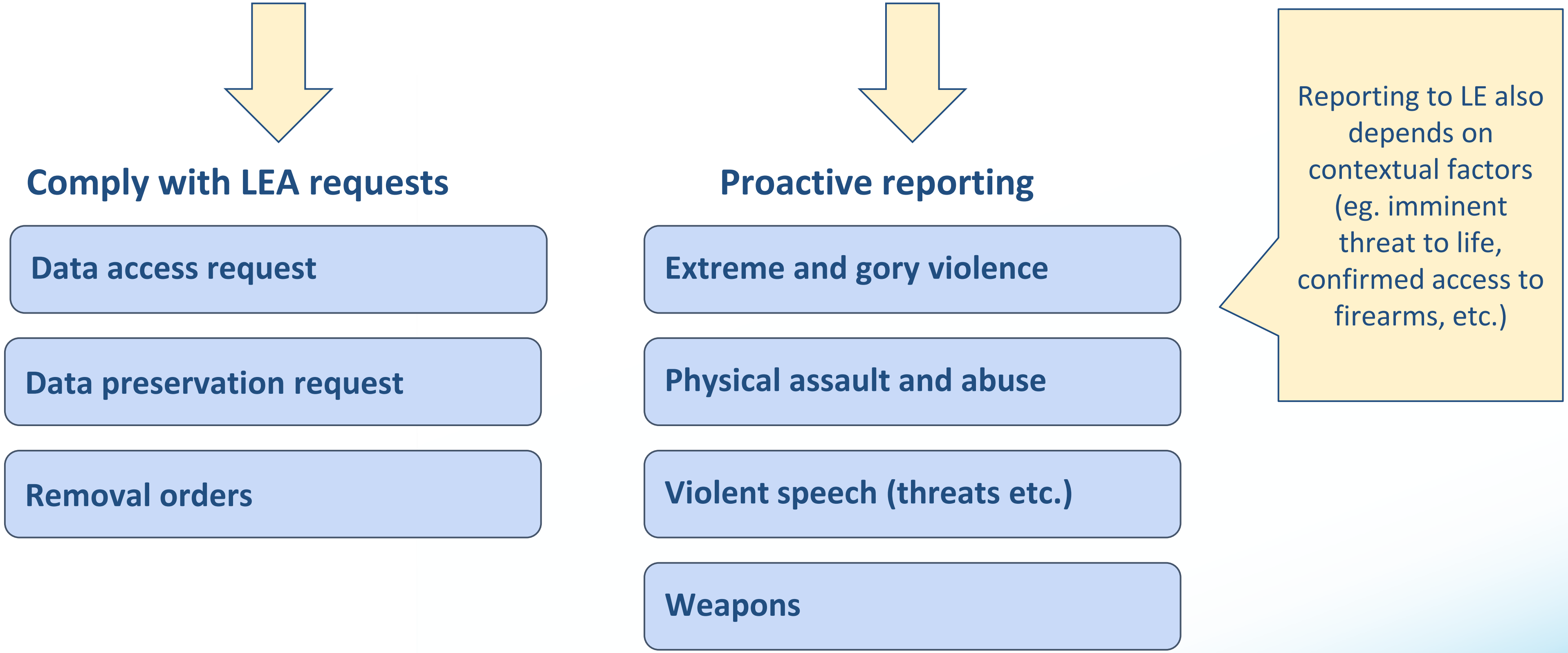
LOGO

network

policy mentoring

Cooperating with law enforcement

We have procedures in place to:



Comply with LEA requests

- **Data access request**
- **Data preservation request**
- **Removal orders**

Proactive reporting

- **Extreme and gory violence**
- **Physical assault and abuse**
- **Violent speech (threats etc.)**
- **Weapons**

Reporting to LE also depends on contextual factors (eg. imminent threat to life, confirmed access to firearms, etc.)

ALLIES technologies and their benefits for HSPs & LEAs

Information Technologies Institute

George Kalpakis

Senior Research Associate at CERTH-IT

ALLIES project

CENTRE FOR RESEARCH & TECHNOLOGY - HELLAS
Information Technologies Institute

MultiMoDal Data Fusion
and Analytics Group

ALLIES technologies & their benefits for HSPs & LEAs

George Kalpakis | Senior Research Associate

CERTH-ITI, Multimodal Data Fusion & Analytics Group

ALLIES Webinar: Terrorist Content Online (TCO): how to prevent it?

11.10.2023

ALLIES target group

CERTH: Centre for Research & Technology HELLAS

Target Group

- **HSPs** (**micro** and **small** service providers, networks)
- **Professionals** (**LEAs**, policy makers, lawyers, prosecutors,)
- **Multiplicators** (networks, civil society organizations, lobbying groups, related projects)
- **General Public**

TCO regulation: Micro and small HSPs

CERTH: Centre for Research & Technology HELLAS

- **TCO Regulation (Regulation (EU) 2021/7845)**
  - HSPs should **remove TCO within one hour,** after receiving an official removal order by a competent authority
  - Enforced to all HSPs regardless of their size and/or resources

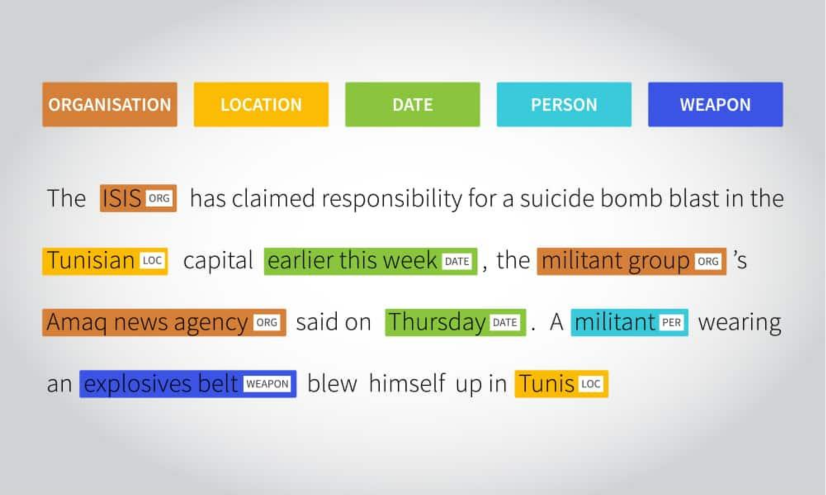
- Limitations for **micro & small HSPs**
  - Limited **human resources**, i.e., <10/50 employees
  - Limited **operational/financial capacity**

ALLIES Technologies

Text Analysis

Pompeu Fabra University

- Processing of **multilingual textual content** hosted online by HSPs
- Automatic **Named Entity** & **Key Concept** recognition
- **Machine translation** & **transliteration**


ORGANISATION
LOCATION
DATE
PERSON
WEAPON

The ISIS ORG has claimed responsibility for a suicide bomb blast in the Tunisian LOC capital earlier this week DATE , the militant group ORG 's Amaq news agency ORG said on Thursday DATE . A militant PER wearing an explosives belt WEAPON blew himself up in Tunis LOC

Speech Recognition

Pompeu Fabra University

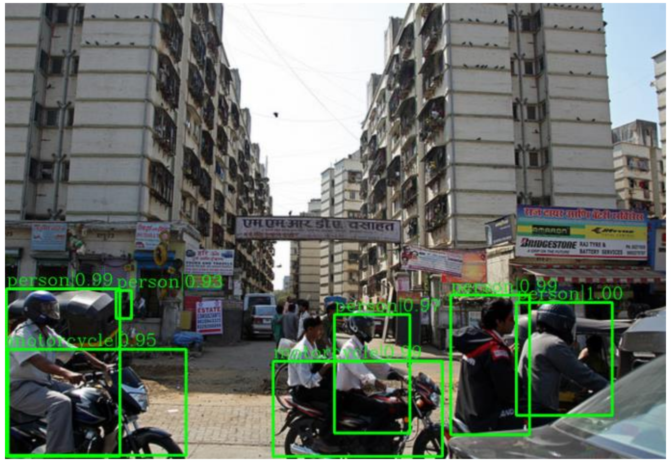
- Multilingual **automatic speech recognition** of the **spoken content** hosted online by HSPs

- Supporting languages of interest for the ALLIES stakeholders

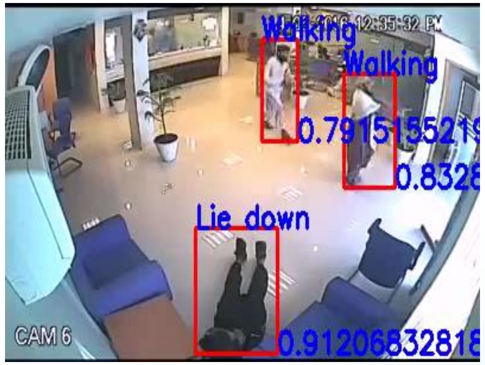
Visual Understanding

CENTRE FOR RESEARCH & TECHNOLOGY, HELLAS

- Analysing **multimedia content** hosted online by HSPs



- **Object recognition** on images/videos
  - Detection of weapons, logos, flags and/or other TCO-related classes



- **Activity/human behaviour recognition** on videos
  - Violent behaviour, bullying, etc.

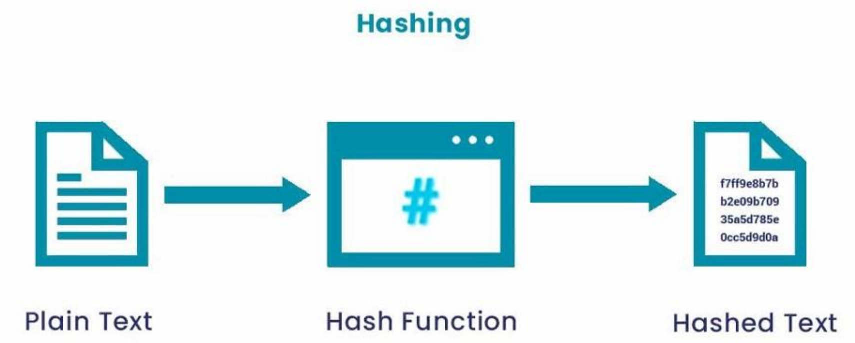
Multimodal Data Hashing

CENTRE FOR RESEARCH & TECHNOLOGY, HELLAS

- Generating **hash representations** of the multimedia content hosted online by HSPs

- **Information retrieval** of duplicate or similar content based on hash representations

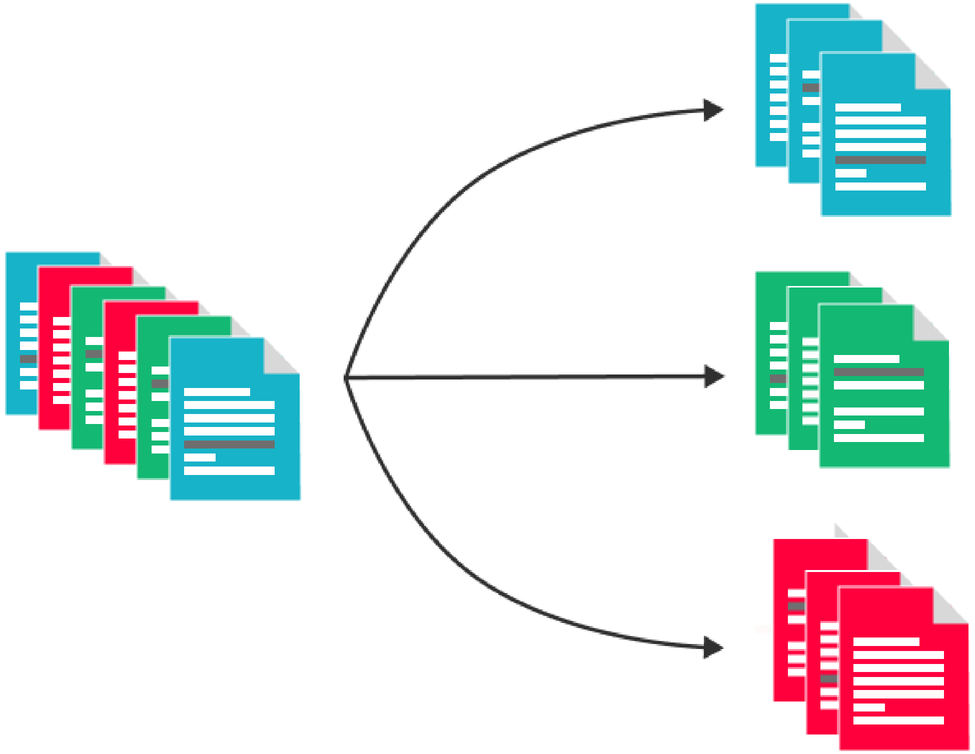
- Modalities: image, video, text
  - **Unimodal** hashes (e.g., image **or** text)
  - **Multimodal** (e.g., image **and** text)



Multimodal Classification

CENTRE FOR RESEARCH & TECHNOLOGY, HELLAS

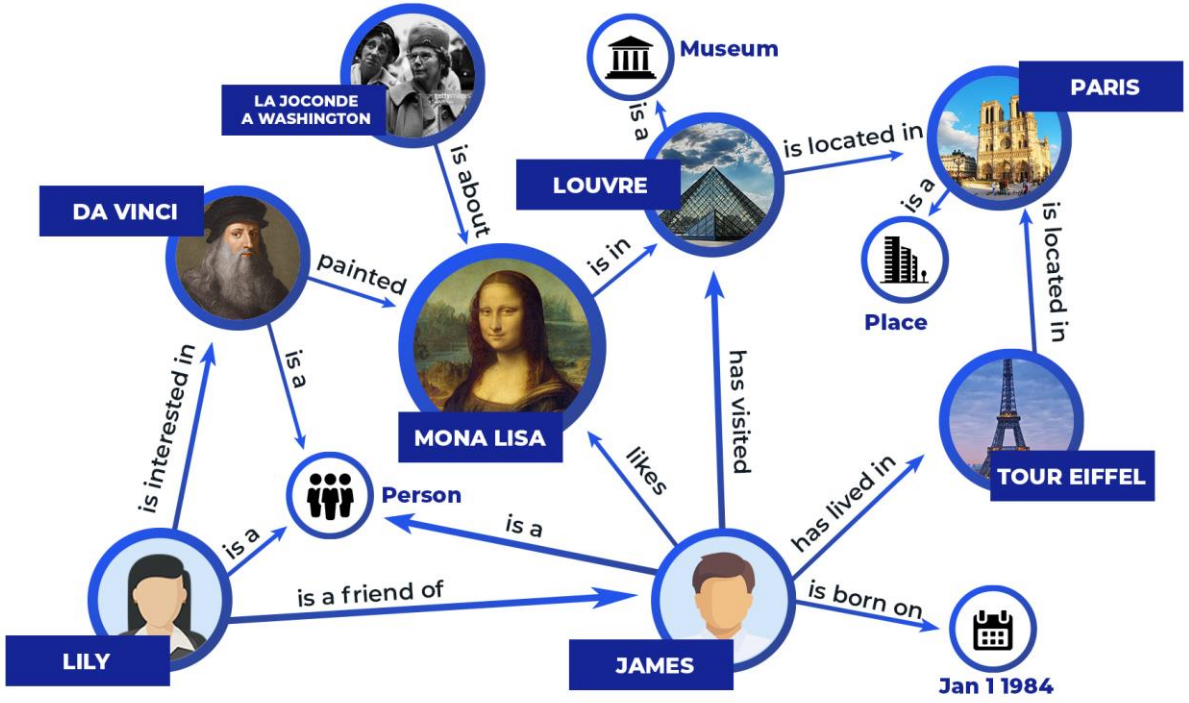
- Categorising **multimodal content** hosted online by HSPs according to its relevance to TCO
- **Modalities**: image, text
- **Training AI models** based on deep neural networks
- **Annotated datasets** provided by the ALLIES end users



Knowledge Graph

novelcore

- **Decision support** based on **Knowledge Graphs**
- Creating **semantic relations** for different resources
- **Information retrieval** by matching specific criteria

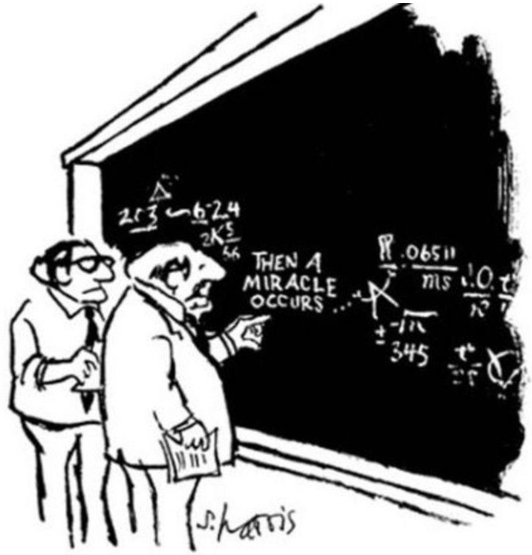


Explanation Engine

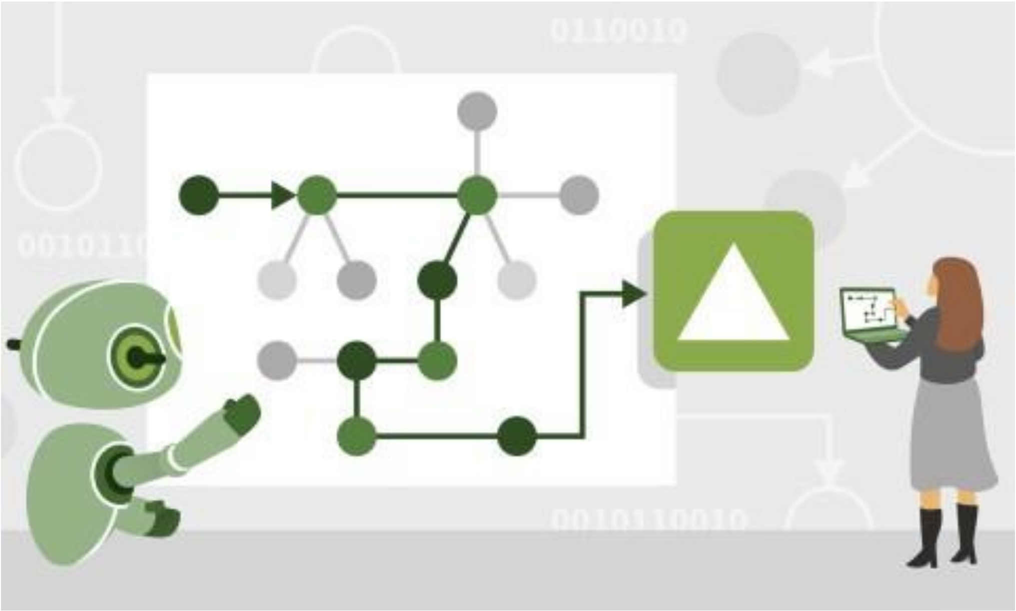
novelcore

- Diverse **explainability** and **reporting** based on AI

- Providing the **reasoning** behind the AI-based outcomes



"I think you should be more explicit here in step two."



Risk Assessment

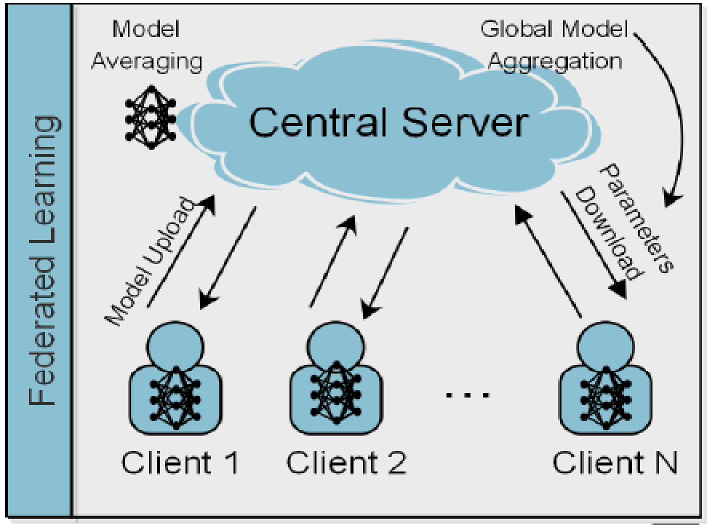
Universita Cattolica del Sacro Cuore

- **Identifying** and **assessing** the **level of risk** of content hosted online by HSPs in terms of TCO
- **Explaining** the relevant outcomes

Federated Learning

Catalink

- **Decentralised** machine learning (ML) approach
  - Raw data **not exchanged** among training nodes
  - ML models **trained locally** & transferred to a central server
  - Ensures **privacy** and **security**: Data never accessed or processed by other parties, i.e., HSPs
  - **Candidate tools**: Visual understanding, Multimodal classification



ALLIES Engagement Platform

Catalink

- **Federated UI**
  - HSP moderators access the outcomes of the analysis for data on their end
  - Outcomes accompanied by comprehensive explanations
- Platform for **removal orders**
  - LEAs submit TCO removal orders in a **standardised manner**
  - HSPs can easily respond/address removal orders
  - Efficient and effective process progress with real-time tracking of the order status
  - HSPs can proactively report TCO
- Shared **hash repository**
  - Populated with hashes of validated TCO
  - Equipped with **hash-comparison** services

Impact & Benefits

Benefits for micro & small HSPs (1)

- 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**

Benefits for micro & small HSPs (2)

- Voluntary **proactive reporting** of TCO

- **Collaborative** AI model **(re-)training** without sharing raw data

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

Benefits for LEAs

- Communication of removal orders across different HSPs in a **standardised manner**

- Capability to monitor the progress of removal orders in real-time

- Receiving reports from HSPs on a voluntary basis

Questions and Discussion

QUESTION LEAs:

How has the TCO Regulation influenced your agency's approach to international cooperation and information sharing in the context of tackling transnational crime, and what mechanisms or initiatives have you established to enhance collaboration with foreign counterparts?

QUESTION HSP:

What measures or technologies is your company exploring or implementing to ensure the security and privacy of customer data while complying with the TCO Regulation's requirements, especially when dealing with sensitive information or international clients?

QUESTION LIF/CERTH:

What are the ethical and legal implications of using AI for monitoring and tackling Terrorist content online? Are ethical and legal implications taken into consideration in the tool development?

ALLIES AMBASSADOR

Let’s collaborate!

Check out the benefits for Hosting Service Provider!

BENEFITS

- Online promotion and visibility on EU level
- Reputation Boost
- First-hand information and invitation to events
- Free participation of TCO educational activities
- Become the first ALLIES AI-tool user for FREE

Let’s collaborate!

Thank you!