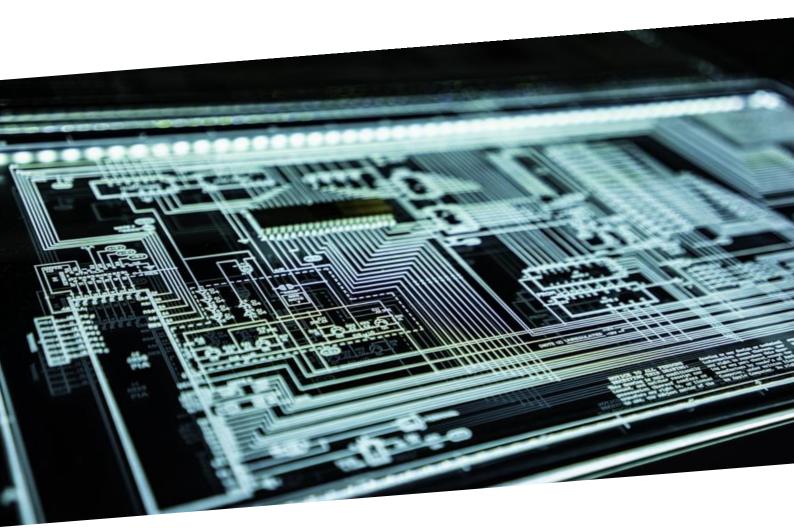


## **BROCHURE**

# Tools & Approaches for small HSPs to address Terrorist Content Online

























#### Fighting teRrorlSt Content Online - FRISCO

FRISCO (2022-2024) is an EU-funded project whose main objective is to raise awareness among micro and small online platforms ("Hosting Service Providers") and to help them comply with the *EU Regulation 2021/784 on addressing the dissemination of Terrorist Content Online* ("TCO Regulation"). The TCO Regulation, applicable as of the 7<sup>th</sup> of June 2022, sets out new obligations for HSPs offering their services in the European Union.

FRISCO is being implemented by 8 partners, from 6 different EU Member States, and directly addresses the dissemination of terrorist content online. It aims to help HSPs comply with the TCO Regulation through tailored training, compliance tools and best practices guidelines. By supporting the fight against terrorist content in Europe, the FRISCO Project is helping to prevent and counter violent extremism online and to create a safer online environment.

#### **About the Partners**

- The National Centre for Scientific Research Demokritos is the largest Greek self-governing research organisation under the supervision of the Greek Government. NCSR - Demokritos takes part in the Project through its Software and Knowledge Engineering Lab at the Institute of Informatics and Telecommunications.
- Tremau is a Paris-based technology start-up, part of the European Trust & Safety ecosystem. It focuses on developing innovative tools to assist actors across the digital industry and mitigate their regulatory and operational risks.
- The International Network Against Cyber Hate is the leading network within the EU and globally that combats cyber hate using a holistic approach that combines monitoring, content removal, educational methods, international cooperation, advocacy, and cuttingedge technology.
- The French Ministry of Interior is represented in the FRISCO Project by several of its units, including Pharos (the French platform for the harmonisation, analysis, cross-checking, and orientation of alerts relating to illicit content online). It is supported by its international cooperation agency, Civipol.



- The European Digital Learning Network ETS (Dlearn) is a non-profit association aimed to embrace the challenges brought by the digital transformation in terms of digital skills mismatch and digital learning opportunities. It is acting in fostering digital learning for every citizen, of any age, at all levels.
- Violence Prevention Network is one of Europe's leading hubs of practitioners, working nationally and internationally with over 120 experts. The NGO's purpose is to support frontline practitioners, policy makers and researchers to bring about positive change in local efforts to prevent and combat violent extremism.
- The ICT Association of Hungary (IVSZ) is the joint platform of the information technology, telecommunications, and electronics sectors, representing the interest of the Hungarian information and communication technology sector for over 30 years with 350 members.

















This project has received funding from the European Commission - Internal Security Fund, under Grant Agreement No 101061643. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.



## **TABLE OF CONTENTS**

INTRODUCTION	5
Presentation of FRISCO	5
About this Brochure	7
FRISCO COMPLIANCE TOOLS (1)	8
Self-Assessment Questionnaire	8
Process Map	8
CONTENT MODERATION SOFTWARE (2)	9
Tremau - France	9
Utopia Analytics - Finland	9
Bodyguard - France	10
DETECTION TOOLS (3)	11
Sightengine - France	11
Textgain - Belgium	11
Identrics - Bulgaria	12
Logically AI inc UK	13
ALLIES Project	14
HASH-MATCHING TOOLS (4)	15
Web-IQ - Netherlands	15
Tech Against Terrorism	15
CONCLUSION	16
CONTACTUS	16



#### INTRODUCTION

#### **Presentation of FRISCO**

The EU Regulation 2021/784 on addressing the dissemination of Terrorist Content Online ("TCO Regulation") provides a comprehensive legal framework at European level to curb the dissemination of terrorist content online and protect European users. It is applicable as of the 7th of June 2022 and sets out new obligations for Hosting Service Providers (HSPs) offering their services in the European Union. HSPs are defined as online platforms storing information provided by and at the request of a user in the EU - i.e., with user-generated content publicly available.

Terrorist content is defined as text, images, audios, or videos used to incite terrorist acts, give instructions on how to commit offences or solicit participation in terrorist groups. Regarding the main measure introduced by the TCO Regulation, HSPs now have one hour to remove content flagged by a 'removal order'. HSPs that do not systematically comply with the rules are liable to be fined up to 4% of their global turnover.

To respond to the challenges met by micro and small HSPs in implementing the TCO Regulation, the European Commission awarded three projects in 2022, including the FRISCO Project. Carried out by a team of 8 partners between 2022 and 2024, the FRISCO Project has one main objective: raise awareness among micro and small HSPs and help them comply with the TCO Regulation.

#### It has been designed to:

 Support HSPs in their compliance and content moderation efforts.

- Alleviate their operational burdens relating to the TCO Regulation.
- Foster multi-stakeholder collaborations and partnerships.

#### **OBJECTIVES**

The FRISCO Project is seeking to raise awareness among micro and small HSPs, foster their compliance with the TCO Regulation, and help them protect their services against terrorist content. Support is provided through:

- Awareness raising informing HSPs about their new obligations.
- Tools production developing tools, frameworks, and mechanisms for them.
- Knowledge-sharing sharing information, experiences, and best practices.

#### **ACTIVITIES**

The FRISCO Project is currently producing three main types of resources to support HSPs in their compliance and content moderation efforts:

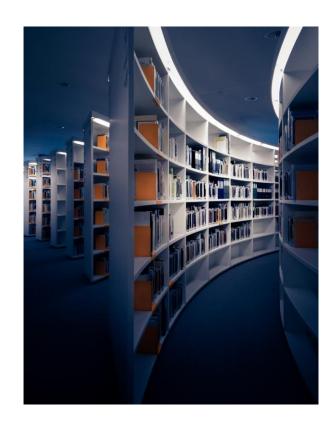
- A Toolbox: a questionnaire, a process map, and a content moderation tool.
- A Training Programme: seven training modules accessible via an online platform.
- Best practices materials: a manual, three workshops and additional resources.
- The FRISCO Project is also engaging in research activities, while disseminating its results and creating partnerships.



#### **INSIGHTS**

In 2023, the FRISCO Project conducted a study (available in full-length here) to assess the readiness of HSPs regarding the TCO Regulation. Following discussions with over 50 HSPs and stakeholders, several challenges were identified:

- Small platforms are heavily targeted by terrorists, but some services are more exposed (i.e., file storage, file sharing, archiving, content pasting).
- Small platforms have a very limited awareness and knowledge of the TCO Regulation but perceive themselves to be at low risk of being exploited by terrorists.
- Small-scale HSPs are currently missing the core infrastructure and the appropriate resources to identify, monitor and remove terrorist content



#### **SERVICES**

To respond to these gaps, the FRISCO Project is offering a lot of services to HSPs and other stakeholders:

- Multi-stakeholder collaborations: we encourage collaboration between HSPs and other stakeholders (e.g. LEAs) to better combat the threat together.
- Knowledge of the legal framework: we offer practical advice and resources to help HSPs better understand the legal framework and the specific steps they need to take to comply with it.

- Free and actionable resources: we develop free resources for HSPs to foster their compliance with the TCO Regulation and counter the terrorist exploitation of the Internet.
- Practical examples: we highlight case studies and real-life examples of practices employed by stakeholders to moderate and deal with terrorist content online.
- Workshops and consultations: we engage with LEAs, HSPs, industry associations, researchers and experts to gather insight, learn from each other and create a safer online environment.



#### **About this Brochure**

The FRISCO Brochure, "Tools & Approaches for small HSPs to address Terrorist Content Online", provides a comprehensive overview of the technical tools and solutions available to HSPs to address the dissemination of terrorist content on their services and increase their compliance with the TCO Regulation. They are presented into the main following categories, which make up the chapters of this brochure:

- FRISCO Compliance Tools: these tools are the ones fully developed by the FRISCO Project and focusing on compliance processes.
- Content Moderation Software: tools offering a full content moderation software including detection tools.
- Detection Tools: tools offering analysis systems for detecting harmful or illegal content using AI systems.
- Hash-matching Tools: software solutions that use cryptographic hash functions to compare digital files or data sets and identify matches or similarities.

The FRISCO Project aims at sharing experiences and best practices to support HSPs in their content moderation and compliance efforts. This brochure responds to this objective by introducing solution providers whose products enable HSPs to simplify and improve their content moderation and compliance processes. It is based on an Online Market Event organised by the FRISCO Project on 18 and 19 April 2024 in Budapest and has also been enriched by insights gathered at our other multi-stakeholder events and during the production of the FRISCO Best Practices Manual.

All the companies listed have been contacted to participate in our Online Marketplace Event where they could pitch their solutions and services to the HSPs invited. Prior to their participation, they were requested to answer several questions (i.e. regarding the services offered, their targets, pricing and licensing, etc.). The answers to these questions were used as the basis for the content of this brochure. Additional information was retrieved by contacting the companies independently. Each company has a section of its own, with a detailed description of its services and solutions.

In short, the FRISCO Brochure has been designed to equip HSPs with a comprehensive list of technical solutions that they can use to refine their content moderation processes and reach full compliance with the TCO Regulation. These technical solutions are offered by a diverse range of service providers, each specialising in various software-based measures to address terrorist and other harmful online content. These span from specific detection tools to all-inclusive content moderation software and advanced hash-matching tools.

All the solution providers presented form an integral part of our global ecosystem, collectively striving to make the internet a safer place for all by preventing the dissemination of harmful and illegal content. This brochure, therefore, serves as a platform for fostering collaboration among various stakeholders. It encourages the creation of synergies and promotes the sharing of experiences, ultimately contributing to a safer digital environment for everyone.

We hope that this document will prove valuable to HSPs, enabling them to find tailored tools that cater to their unique needs within the industry.

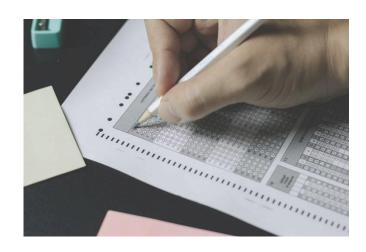


## FRISCO COMPLIANCE TOOLS (1)

#### **Self-Assessment Questionnaire**

The self-assessment questionnaire aims at helping small and medium HSPs understand their level of compliance with the TCO Regulation. This questionnaire is meant as a first step for HSPs to understand how their current internal processes for content moderation regarding terrorist content align with the TCO Regulation. The objective of this tool is to provide HSP with a compliance score, which helps them situate themselves in the path to full TCO compliance.

Available here





#### **Process Map**

FRISCO's process map is an interactive tool that structures and describes the entire compliance process with the TCO Regulation and related duties for HSPs in a holistic way, so to say from the exposure to terrorist content to the transparency reports. This tool, focused on HSPs' operational needs, provides a precise breakdown of the TCO Regulation, step by step, and is based on a and chronological approach compliance. The process is displayed gradually, thanks to interactive and/or animated transitions, and the whole map can be downloaded entirely in the end. Legend, colour code and steps increase ergonomics and user's experience. Finally, more information, resources and tools are made available via the tool.

Available here



## **CONTENT MODERATION SOFTWARE (2)**

#### Tremau - France

friscoproject.eu/content-moderation-tool/

The FRISCO Content Moderation Tool addresses user-generated content related risks. It is based on Tremau's solutions and tailored to HSPs' needs in relation to the TCO Regulation, thanks to FRISCO's findings and resources. As a small HSP, you face unique challenges in ensuring that your platform remains free from any form of harmful content, including terrorist-related materials. Our purpose-built Content Moderation Tool is here to assist you in this critical mission by moderating all content, no matter the source of detection, in one place; streamlining their moderation process; generating operational efficiencies (of up to 50%); achieving better control and transparency; ensuring consistent application of their Terms of Service and Community Guidelines and customising queue management and prioritisation rules.

#### Key features include:

- o All data signals in one place
- Al powered end-to-end moderation platform integrating all detection sources
- Manual and automated processing
- o Compliant by design
- Built to power best practices

If you're interested in a demo or have any questions, please feel free to reach out to us at info@friscoproject.eu or info@tremau.com directly. Our team is ready to provide you with more information and guide you through the capabilities of the Content Moderation Tool.

#### Utopia Analytics - Finland

utopiaanalytics.com

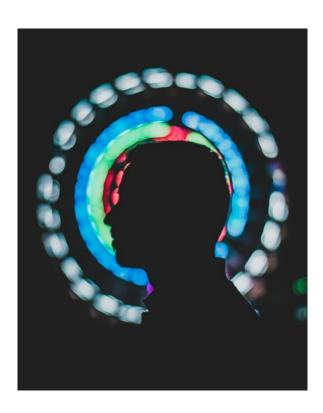
Utopia Analytics is a Finland based company serving different industries since 2014, that automated enables moderation with proprietary AI to identify inappropriate content or behaviour, hate speech, discrimination, sexual harassment, spam, etc. Online content is automatically filtered out 24/7 to allow teams to focus on moderation policy management. Utopia Al Moderator is a fully automated moderation tool that protects your online community and your brand from abusive user-generated content, fraud, hate-speech and illegal content. The AI learns from previous publishing decisions made by your team, works in real-time, with a higher degree of accuracy than traditional tools and it is ready for production in only 2 weeks.

- Rulemaking: the user can develop its moderation policies. The system subsequently learns from moderator decision-making to apply these rules.
- Functional features: Utopia AI moderator can define the reasons why a particular content was flagged.
- Social media language: Utopia AI moderator understands the semantic meaning of social media language regardless of spelling errors, informal grammar uses or unusual meanings of words.
- Image and video moderation: Utopia Al analytics also provides automated video and image moderation.



Utopia Analytics works with leading organisations in their fields, collaborating across several different industries such as: online platforms, news media companies, social media platforms, online marketplaces, gaming and dating companies, reviews sites, among others. This tool can help overcome the challenges related to user-generated content moderation while implementing solutions that help you comply with legislations and regulations such as Digital Services Act (DSA).

**Pricing information**: Utopia Analytics services are offered as Software as a Service (SaaS) with very competitive prices based on the requirements of the project. For more information contact utopiaanalytics.com



### Bodyguard - France

#### bodyguard.ai/en

Bodyguard.ai helps businesses from all industries in curating their user-generated content in multiple languages and across multiple platforms and social media in real-time. Bodyguard's content moderation software has an advanced AI and NLP technology, provides the best contextual analysis of the market, allowing clients to better understand their audience and make informed decisions to improve their social media strategy.

#### Key features include:

- Dedicated dashboard with valuable insights and actionable data
- Context-aware analysis of user-generated content
- Sentiment analysis
- Special Sport and Gaming applicability.

By filtering out irrelevant content and identifying valuable user-generated content organisations can help promote their brand and engage with their customers safely. Thanks to real-time analysis, this tool can respond quickly to potential crises and ensure that their online presence remains positive and engaging.

**Pricing information**: Please contact Bodyguard using their contact form.



## **DETECTION TOOLS (3)**

#### Sightengine - France

#### sightengine.com

Sightengine offers **Automated** Content Moderation models and tools to analyse images and videos based on Natural Language Analysis and Proprietary Algorithms. These tools come via API integration and offer 60+ moderation categories across nudity, hate, violence, drugs, weapons, self-harm and more that are context aware. These tools are highly customizable and can be tailored to clients' needs. It is available in many languages to support users from any region and can be used on messages, chats, user profiles, usernames, comments or reviews in real-time. Whenever there is a match, the Sightengine API gives a tag describing the type of content found along with its position in the text, so the user can easily remove the offending phrase, change it or block the entire message.

#### Key features include:

- Choice of the filters to apply Prevention of users bypassing these filters by detecting millions of variations for each unwanted word: letter replacements, character insertions, phonetic variations, leet speak, obfuscation, hyphenation, special characters, embeddings etc.
- Offers the possibility to decide when users should be allowed to share links and web URLs by implementing a whitelist/blacklist of specific URLs and domains.

**Pricing information**: Sightengine offers different pricing plans depending on the specific needs with a free version available. See the full information here.

#### Textgain - Belgium

#### textgain.com

Textgain was founded in 2015 as a spin-off of the University of Antwerp and offers a full range of customizable data management and analysis solutions to sort incoming emails, extract insights from PDFs, follow the news, or discover opinions on social media. They specialise in the development of AI that automatically detects and monitors harmful societal trends and tensions, such as hate speech or disinformation.

- Text Analytics API offers a collection of diverse NLP tools to help infer insights from text data: age detection, sentiment analysis, gender prediction, language-specific grammar and lexical analytic features, polarity calculation, etc. This product is multilingual (available in 15 languages), without server-side data retention.
- SentiMeter is a data analytics dashboard that uses Textgain API to measure online sentiment among citizens of a specific geographical area, like a city, municipality or province. It is customizable for social media platforms or websites to be monitored, topics, locations, etc. Sentimeter isolates social media data for this area based on geolocation, trigger words and local public accounts. Various NLP strategies are applied on the collected data to perform sentiment analysis and detect trends and shifts on social media. As SentiMeter collects more data, it provides an increasingly accurate representation.



- Grasp is a free, lightweight AI toolkit for Python, with tools for data mining, NLP, machine learning and network analysis. Grasp is a multifunctional toolkit with tools for data mining, including downloading and processing html, finding ready-to-process Wikipedia articles and twitter opinions, as well as easily deploying API's. The natural language processing toolkit includes tools for identification, language finding sentences, and word types, as well as determining the polarity of a text. Grasp's network analysis toolkit includes seamless functions to map and visualise networks. The machine learning toolkit includes algorithms to quantify text, train models, and retrieve predictions. Machine Learning algorithms learn by example. By showing them 10K spam and 10K real emails (i.e., train a model), they can predict whether other emails are also spam or not.
- Panorama is an award-winning explainable Al framework for real-time, multilingual social media analysis. It includes a dedicated annotation app, **GDPR-compliant** anonymization algorithms, anonymous user profiling (age, gender, etc.), named entity recognition, and a stunning visualisation toolkit (charts, graphs, trees, etc.). It detects toxic content in all EU languages (and Arabic, Turkish, Russian) on dozens of online platforms, with fine-grained labelling of ridiculing, dehumanising, discriminatory, misinformative, violent and/or extremist content, carefully curated by 50+ local language experts. Most of all, it is a very lightweight program that does not require heavy hardware. The Panorama package also contains a visualisation component to help the user make sense of their data.

In 2021, Textgain became the coordinator of the European Observatory of Online Hate, an initiative to monitor online hate speech across the European Union.

**Pricing information**: Pricing will depend on the tool. For more information, please contact info@textgain.com.

#### Identrics - Bulgaria

#### identrics.ai

Identrics is an award-winning research and development company that creates AI-based solutions for publishing, media and risk intelligence - from ready-to-use models to fully tailored data science services. These AI-based solutions help organisations of all sizes and the people in them, deal with immense volumes of data and text. Our algorithms delegate organisations' menial data and text processing tasks to machines, so that humans can focus on high-end, insight-oriented work. Solutions range from data classification and indexing to content enrichment – such as named entity recognition, sentiment analysis, all the way to automated content recommendation and generation.

- Natural Language Processing Enrichments:

   a service that enhances text data with contextual information, enabling better understanding and analysis of unstructured data.
- Topic Engagement: a tool that analyses user interactions with specific topics, providing insights into audience engagement and interest levels.



- Hate Speech Detection: a service that automatically identifies and flags hate speech in user-generated content, helping to maintain a safe online environment.
- Propaganda and Disinformation Detection:

   a tool that detects and categorises
   misleading or false information, preventing its
   spread and mitigating its impact.
- Abstractive Text Summarisation: a service that generates summaries of lengthy texts, providing a quick overview of key information without losing the original context.
- Automated Reports and Newsletters: a tool that streamlines the creation and distribution of reports and newsletters, saving time and resources while ensuring consistent and accurate communication.
- Media Contacts Database Automation: a service that automates the process of building and maintaining a database of media contacts, enabling efficient and targeted outreach.
- Custom & DSaaS: customised data science solutions and Data Science as a Service (DSaaS) offerings that cater to specific business needs, providing tailored insights and analytics.

Identrics' disinformation detection services cater to the needs of a diverse audience, including research centres, law enforcement agencies, educational institutions, government bodies, media organisations, fact-checking organisations, and HSPs.

**Pricing information**: For information about the pricing, please use the contact form here.

#### Logically Al inc. - UK

#### logically.ai/

Logically is a tech company identifying malign influence operations and harmful content on the internet. Logically has produced a U.S. Domestic Extremism Content Classifier, Founded in 2017, working with aovernments and now organisations around the world. Logically's Domestic Extremism Content Classifier is an Al model that can identify pieces of online, textbased content within our software solution called, Logically Intelligence, that represent indicators of U.S. domestic extremism. The model specifically focuses on identifying content that could represent hate speech, indoctrination, radicalization, recruitment, or plotting related to anti-LGBTQIA+, racially/ethnically motivated, or anti-government authority violent extremism. While our classifier is up and running on U.S. domestic extremism, the process and methodology we used is repeatable across cultures and languages.

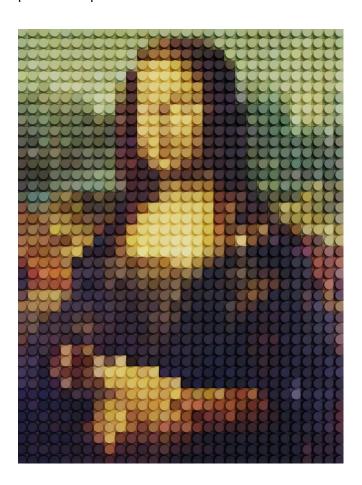
- Counter Narrative Strategies
- Event Influence Assessment
- Group Actor Analysis
- o Influence Campaign Attribution
- Misinformation and Disinformation Analysis
- Policy and Strategy Formulation
- Trend Analysis for Fact-Checking
- Adaptable to different languages



#### **Pricing information.** There are three offering:

- o Purchase of the domestic extremism classifier.
- Purchase of Logically intelligence that includes the U.S. domestic extremism classifier and other threat classifiers.
- Technical services work to design and build an extremism classifier tailored to a different language and culture.

For access to the Logically Intelligence platform's U.S. domestic terrorism product is approximately \$5,000 per month. To develop terrorism threat classifier to match a regions terrorism problem requires a discussion with the DevOps team to provide a quote.



#### **ALLIES Project**

#### alliesproject.com/

The ALLIES Project aims to support micro and small HSPs in reaching full compliance with the new requirements and obligations of the TCO Regulation. This is achieved through the creation of proper learning, training, experience sharing mechanisms as well as technical AI tools, specifically designed and tailored for HSPs.

#### Key feature include:

- Text analysis
- o Multilingual speech recognition
- o Video and Image recognition
- o Multimodal data hashing
- o Risk assessment for content

**Pricing information**: The tools are free to use.



## **HASH-MATCHING TOOLS (4)**

Hash-matching is a fingerprinting technology that allows the creation of a hash for a piece of content (i.e. image, text audio, video) and then compares it to existing database of prohibited hashes allowing to flag illegal content.

#### Web-IQ - Netherlands

#### web-iq.com

Web-IQ is an IT company based in the Netherlands, using Open-Source Intelligence (OSINT) to help organisations fight serious crimes, like fraud, child abuse and human trafficking. Their Hash Check Service transforms any image or video into a hash, which is the digital fingerprint of a file in the form of a code, to its central database of known child sexual abuse material. The user receives an answer as to whether this hash occurs on lists of hashes provided by e.g. the Dutch National Police. This service is provided through an HTTP based API. The Hash Check Service is a service offered by the Dutch hotline Offlimits and Web-IO. The Hash Check Service is meant for Internet Service Providers (ISP's), hosters, and Trust & Safety organisations. The purpose behind the Hash Check Service is to increase the speed with which online child sexual abuse material can be detected and removed from the internet. With the Hash Check Service HSPs can tap into many databases of known child sexual abuse material with an easy integration. In those efforts, Web-IQ has been collaborating with Tax Authorities, Financial Services and Enforcement Law agencies for over 10 years.

**Pricing information**: Any ISP or hoster from a country that is part of the INHOPE organisation can use the Hash Check Service for free.

#### **Tech Against Terrorism**

#### techagainstterrorism.org/home

Tech Against Terrorism is an independent, public-private partnership launched by the United Nations in 2016. They work with the tech sector and are supported by the UN as well as by international bodies including governments of the United Kingdom, Spain, Switzerland, the Netherlands, the Republic of Korea, and Canada. Their work revolves around threat intelligence, technical innovation, and policy advice. Tech Against Terrorism offers a Terrorist Content Analytics Platform (TCAP), a content moderation tool for tech platforms, enabling the detection and removal of terrorist content through open-source intelligence and connected to a Hash-Matching database.

#### Key features include:

- Content reporting and moderation
- Automated content alert
- Automated status checks to determine if collected content is still online as well as automated archiving and hashing of content.

Additionally, Tech Against Terrorism offers TCO training material and a knowledge sharing platform and strives to empower and educate tech platforms to improve their existing counterterrorism responses.

#### **Pricing information**:

contact@techagainstterrorism.org



#### CONCLUSION

As part of the FRISCO Project, this document was created to provide HSPs access to a variety of solutions that can help them with TCO compliance and fight back against terrorist content online. The FRISCO Brochure overviews a total of 12 tools from various service providers. These tools are divided into 4 categories: (1) FRISCO Compliance Tools, (2) Content Moderation Software, (3) Detection Tools, (4) Hash-matching Tools. Through these 4 categories, we hope to cover enough grounds to allow HSPs to find tools tailored to their specific needs and facilitate TCO compliance as well contribute to emerging collaboration in online anti-terrorist activities.

#### **CONTACT US**

FRISCO is an EU-funded project implemented by a team of 8 partners across Europe. Our main objective is to raise awareness among small tech companies and online platforms and to help them comply with the TCO Regulation. By supporting the fight against terrorist content in Europe, we are helping to prevent and counter violent extremism online and to create a safer online environment. If you wish to contact the FRISCO Project, use the following e-mail address: info@friscoproject.eu.





















