tech against terrorism europe

Improving Your Counter-Terrorism Response

A Six-Step Guide to Adhering to the EU's TCO Regulation

Sophia Rothut, Heidi Schulze, Diana Rieger, Catherine Bouko and Brigitte Naderer

Funded by the European Union

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

GHENT UNIVERSITY

Tech Against Terrorism Europe (TATE)

This guide is part of Tech Against Terrorism Europe (TATE). The TATE project is funded by the European Union (ISF-2021-AG-TCO-101080101). This project supports smaller hosting services providers (HSPs) in building their counterterrorism frameworks and with transparency reporting, as required by the EU's terrorist content online (TCO) regulation and in Directive (EU) 2017/541.

tate.techagainstterrorism.org

Funded by the European Union

Authors

Sophia Rothut is a Research Associate in the lab of Professor Rieger (LMU Munich). As part of the EU project TATE, she is working on requirements for countering terrorist content online. Her research focuses on online radicalisation, mainstreaming of radical ideas, and political/far-right influencers.

Heidi Schulze is a Research Associate at the Department of Media and Communication at LMU Munich. At LMU, she is part of the Research Lab of Professor Rieger and studies radicalisation dynamics online within the large-scale research project MOTRA – Monitoring System and Transfer Platform Radicalisation. In her research, she focuses on radical/extremist (group) communication in alternative social platforms and fringe communities, as well as characteristics and audiences of hyperpartisan news websites.

Diana Rieger is Full Professor at the Department of Media and Communication at LMU Munich. She conducts research on online radicalisation, hate speech, and the effects of entertainment content. She has also published on the development and evaluation of countermeasures against radicalisation.

Catherine Bouko is Associate Professor of Communication and French at Ghent University (Belgium). Her main research is on political communication, extremism, and citizenship on social media, with a special focus on image-based communication. Methodologically, she employs (multimodal) discourse analysis, quantitative content analysis, semiotics and ethnography.

Brigitte Naderer is a Post-Doctoral Research Associate in the Center for Public Health, Department of Social and Preventive Medicine, Suicide Research and Mental Health Promotion Unit at the Medical University of Vienna. She previously worked at LMU Munich's Department of Media and Communication until March 2023. Her research focuses on media literacy, online radicalisation, and media effects on children and adolescents.

Acknowledgement: We thank Tech Against Terrorism for their input and feedback on this guide. Also, we would like to thank Professor Maura Conway for her helpful comments.

tech against terrorism europe

Table of Contents

A. Introduction

What this guide is about, and why you should read it…

This guide deals with the **obligations placed on hosting service providers (HSPs) by the European Regulation on Terrorist Content Online (TCO)** adopted in April 2021. It addresses what HSPs need to consider to counter the dissemination of terrorist content online and gives practical advice on how to implement various measures for doing so.

The guide focuses firstly on the **minimum requirements HSPs must satisfy to comply with the TCO Regulation**. Secondly, it provides **tips and practical advice on relevant (pro-)active measures HSPs should adopt** to successfully navigate the complexities of regulatory enforcement (i.e. what to do when a removal order appears in your mailbox) and to prepare HSPs to counter terrorist exploitation of their platforms.

The guide is part of the Tech Against Terrorism Europe (TATE) project funded by the European Commission. The TATE consortium composes seven partners: Dublin City University, Ghent University, JOS Project, LMU Munich, Saher Europe, Swansea University, and Tech Against Terrorism. TATE aims to create awareness of the EU TCO Regulation and to support small and micro tech companies to take measures in support of it. Together with other resources, TATE has created this guide to make the legal and practical expectations as tangible as possible for HSPs.

Who is this guide aimed at?

The guide is aimed at hosting service providers (HSPs) and their employees as well as at IT professionals who want to implement technical features for countering terrorist content in platform architectures. It is intended to provide information about the minimum requirements that must be satisfied to comply with the EU's TCO Regulation.

We have created this guide and other TATE resources to support small and micro HSPs. We acknowledge that small and micro HSPs often have limited resources to tackle terrorist content on their platforms. **However, it is crucial that small and micro HSPs do not neglect the threat since smaller platforms are more likely to be exploited by terrorist actors** (for details, see this Tech Against Terrorism report).

What are the core elements of the EU Regulation on Terrorist Content Online (TCO)?

The European Regulation on Terrorist Content Online (TCO) came into effect in June 2022 and **obliges HSPs to remove content or disable access to it within one hour of receiving a removal order from a competent authority**.

HSPs have to cooperate with law enforcement authorities, like Europol, as well as other relevant authorities, in detecting and removing terrorist content that may be present on their platforms. To comply with the TCO Regulation, **HSPs must also implement effective and proportionate measures**

to prevent terrorist content from being re-uploaded and face further obligations that you will learn about in this guide.

How is this guide structured?

The TCO Regulation currently requires HSPs to:

1) Draft appropriate Terms of Services (▶chapter 1),
2) Take specific measures for identifying and removing terrorist content (▶chapter 2),
3) Establish effective moderation mechanisms (▶chapter 3),
4) Establish points of contact and legal representatives (▶chapter 4),
5) Set up user notification and complaint mechanisms (▶chapter 5),
6) Publish transparency reports (▶chapter 6).

We have structured the explanations and recommendations provided in this guide according to these six principal requirements. At the beginning of each chapter, you will find a short summary of the chapter's contents and its main points. Thereafter, you will be provided with detailed information on what the TCO Regulation requires or urges HSPs to implement as well as further practical advice to defend your platform against terrorist (and other harmful) content.

Which platforms are affected by the TCO Regulation?

The TCO Regulation affects hosting service providers, which includes any platform enabling users to distribute information to the public via its services.

The TCO Regulation applies to **HSPs of all sizes and to every HSP offering its services in the EU**. It also applies to HSPs located outside of the EU when a HSP (1) has a significant number of users in one or more EU Member States or (2) targets its activities at one or more EU Member States.

When is a HSP regarded as 'exposed to terrorist content'?

The TCO Regulation places specific obligations on HSPs that are 'exposed to terrorist content'. According to Art. 5.4, such exposure occurs when a HSP has been notified of and received two or more final removal orders in the previous 12 months from the Member State competent authority in which the HSP has its main establishment or its legal representative in the EU.

What is terrorist content?

Since the TCO Regulation is all about terrorist content disseminated online, it is key to provide a definition. EU Directive 2017/541 lays the basis for the TCO Regulation by defining what is understood as terrorist content.

Content is regarded as terrorist when it incites to commit actions or fosters intentions in favour of a terrorist cause, thus directly or indirectly contributing to the threat of terrorist offences.

Threatening to commit a terrorist offence also counts as terrorist content, as does the provision of information, support, or funding for these acts.

Types of terrorist offences

Terrorist offences can include the following (EU Directive 2017/541, Art. 3.1):

- Attacks upon a person's life or physical integrity
- Kidnapping or hostage-taking
- Causing extensive destruction to specific facilities and infrastructure (e.g. governmental/ public facilities, transport and information systems) likely to endanger human life or result in major economic loss
- Seizure of aircraft, ships or other means of public or goods transport
- Manufacture, possession, acquisition, transport, supply or use of explosives or weapons and research into, and development of, chemical, biological, radiological or nuclear weapons
- Release of dangerous substances, or causing fires, floods or explosions, and the disruption of fundamental resources (e.g. water, power) by which human life is endangered
- Interfering with or disrupting the supply of water, power or any other fundamental natural resource, by which human life is endangered.

Directing a terrorist group or intentionally participating in its activities is also regarded as a criminal offence (EU Directive 2017/541, Art. 4). This includes supplying (informational) resources (e.g. instructions or materials to build weapons) for terrorist causes or funding of such activities.

With such offences, terrorist actors aim to (a) seriously intimidate the public, (b) unduly compel a government or an international organisation to (abstain to) carry out a specific act, or (c) seriously destabilise or destroy the fundamental political, constitutional, economic or social structures of a country or an international organisation (EU Directive 2017/541, Art. 3.2).

To sum up, content is considered to be terrorist content when it enables, supports, or facilitates a terrorist offence or when a threat to commit a terrorist offence is included in it.

B. Key Obligations and Recommendations in Relation to the TCO Regulation

In the following chapters, you will find information about six key components of the TCO Regulation and steps to harden your platform against the terrorist threat. These include the following key areas and questions.

Chapter 1: Drafting and Applying Terms of Service Prohibiting Terrorist Content

1. What are Terms of Service (ToS)?
2. Why is it necessary to have clear and robust ToS?
3. Practical tips and advice: What are elements of robust ToS?

Chapter 2: Specific Measures for Identifying and Removing (Terrorist) Content

1. Why is it necessary to establish identification processes for illegal and harmful content?
2. A process for identifying terrorist content
3. Practical tips and advice: What helps to assess whether the content is illegal?
4. What should I do if I see it differently? How to challenge a received removal order

Chapter 3: Establishing Effective Moderation Mechanisms for Terrorist Content Online

1. What is content moderation and why is it necessary in some cases?
2. Practical tips and advice: How should content moderation be implemented?
3. Alternative moderation approaches

Chapter 4: Establishing Points of Contact and Legal Representatives

1. What are contact points and legal representatives?
2. Why is it necessary to have a contact point or legal representative?
3. What is the competent authority of an EU Member State and how do I contact them?

Chapter 5: Setting Up a User Notification and Complaint System for Removed Content

1. Why is it necessary to set up a transparent complaint mechanism?
2. What are the requirements for a complaint system?
3. How are complaints to be handled and what are the possible results?
4. Practical tips and advice: What elements are useful when establishing a complaint system?

Chapter 6: Practical Support and Advice Around Transparency Reporting

1. What are transparency reports?
2. Why are transparency reports necessary?
3. A process for the preparation of transparency reports
4. What information and metrics need to be included in the TCO transparency report?

The guide explains the **aspects of the TCO Regulation that are most relevant for platforms while also providing practical advice on proactive measures that can be taken to counter the spread of harmful content online**. Such measures are essential for preparing platforms against terrorist exploitation of their services – or in the language of the TCO Regulation: to be able to handle the situation when the first and any subsequent removal order lands in your mailbox.

Chapter 1

Drafting and Applying Terms of Service Prohibiting Terrorist Content

Summary: Contents and Main Points of this Chapter

- Terms of Service (ToS) are a **binding agreement** between the user and the HSP, defining appropriate and permitted platform use.
- HSPs shall (1) set out their **strategy to address the dissemination of terrorist content** in their ToS and (2) **prohibit** the dissemination of **terrorist content**.
- Furthermore and going beyond the scope of the TCO Regulation, HSPs can and should consider prohibiting other forms of harmful content (e.g. extremist content, hate speech).
- ToS are a **legal necessity** under the TCO Regulation and can also build in **useful protections for a platform**.
- Several elements are required to make your ToS as robust as possible, including a definition of terrorist content and the disclosure of the HSP's strategy to combat terrorist content on its platform.

Appropriate **Terms of Service (ToS) are the basis for prohibiting and, consequently, handling terrorist content.** The TCO Regulation explicitly urges HSPs to set out "their policy for addressing the dissemination of terrorist content, including, where appropriate, a meaningful explanation of the functioning of specific measures, including, where applicable, the use of automated tools" (TCO Regulation, Art. 7.1).

Additionally, and going beyond the scope of the TCO Regulation, HSPs can and should **consider prohibiting other forms of harmful content** (e.g. extremist content, hate speech, advocacy of violence) in their ToS. In doing so, HSPs can contribute to the creation of a framework for a civil digital culture.

1. What are Terms of Service (ToS)?

ToS are rules set by and for specific platforms that define (1) HSPs' responsibilities to their users, and (2) the appropriate and permitted, but also prohibited conduct and content on that platform. Users have to accept these terms if they want to use HSPs' services.

Synonyms used for ToS are, for instance, terms of use, terms and conditions, or community standards. The TCO uses the phrase 'terms and conditions' and defines it as "all terms, conditions

and clauses, irrespective of their name or form, which govern the contractual relationship between a hosting service provider and its users" (TCO Regulation, Art. 2.8).

2. Why is it necessary to have clear and robust ToS?

For a deeper look into why it is necessary to have clear and robust ToS, we introduce two different perspectives: the legal and the corporate/operational.

a) Legal perspective

Compliance with (EU) law

ToS are required to comply with current (EU) law, including the TCO Regulation but also EU Directive 2017/541, which strengthens EU-wide handling of terrorist content by (1) providing a definition of terrorist content, (2) setting penalties for it (Art. 15), (3) strengthening the victims' rights and support, and (4) acknowledging terrorism as a transnational, cross-border threat. Both the TCO Regulation and the Directive foster EU-wide and international cooperation. ToS are a venue where HSPs can emphasise their commitment to countering terrorist activity.

Protection of the HSP from legal liability

ToS are intended to protect a HSP from liability (as long as ToS are conform with applicable law). Since it is a contract between the HSP and its users, ToS are a legally binding agreement between these two parties. ToS allow the HSP to set out rules for (non-)acceptable use in accordance with the HSP's values.

Justification of (proactive) content removal

Furthermore, clear ToS are important when encountering and moderating problematic content. HSPs can refer back to their ToS to justify content moderation if they include detailed information on prohibited content practices, including a prohibition of terrorism.

b) Company perspective

Upkeep of the operational business

First, it is central to the HSP's operational business and, thus, to the success of the company to comply with legislation. HSPs have a certain responsibility towards their shareholders, who expect compliance. In addition, financial liability may be incurred, and regulatory penalties or reputational damage are further possible consequences of non-compliance.

Demonstration of HSPs' civic responsibility

Second, as prominent societal figures, HSPs have a responsibilities to the public, both collectively and individually. The responsibility to the public at large involves countering illegal activities such as terrorism and detecting and preventing terrorist content before it is widely disseminated. By stating in the ToS that illegal activities are prohibited, HSPs lay a foundation for a safer online sphere and can rely on it when actioning illegal content. Clear ToS also help

companies demonstrate their responsibility to individual members of the public by building trust through transparency. In this respect, HSPs demonstrate to individual users that they will be protected from harmful content on the platform when such content is explicitly and enforceably prohibited in their ToS.

3. Practical tips and advice: What are elements of robust ToS?

The following sections provide guidance on elements that you, as a HSP, should consider when drafting ToS in the context of the TCO Regulation.

It is important to emphasise that developing ToS is an iterative process: ToS can and should be adapted if necessary, for example, in response to user suggestions and by keeping pace with national or international legislative developments. The TCO Regulation calls for such adaptability: HSPs, if they are exposed to terrorist content, are required to amend their ToS to state that and **outline what actions will be taken to counter misuse of the platform for terrorist purposes** (TCO Regulation, Art. 5.1).

Specifically, we recommend paying heed to the **clarity and structure** of the ToS. Studies show that users spend little time reading ToS and do not pay sufficient attention to the the information presented; ToS rather act as a remider to users' general perception of what is allowed and what is not[1,2]. To aid navigation and comprehension of the ToS, it is advisable to invest time in the visual design of ToS and to work with clear headings or bullet point lists. It can also be useful to remind users of ToS aspects from time to time and, if suspicious behaviour has been detected, to warn of possible consequences (e.g. via a pop-up)[2].

When setting up and revising the ToS, the following **checklist can serve as a guide to optimise compliance with the TCO Regulation**.

Define terrorist content

Having working definitions of terrorism and terrorist content is important for actioning content and behaviour against. These definitions should be appear in the ToS and be referred to when evaluating content. Existing definitions already presented and discussed in this guide (►see the definition here), in particular those of the EU based on EU Directive 2017/541, can provide useful guidance in this regard.

Disclose your strategy for combating terrorist content

The TCO Regulation obliges HSPs exposed to terrorist content to explain in their ToS both their strategy for combating the dissemination of such content and any automated means used (e.g. in identifying prohibited content; ►chapter 2). The TCO Regulation further requires HSPs to implement specific measures once exposed to terrorist content (e.g. systems for

[1] Obar, J. A., & Oeldorf-Hirsch, A. The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services. *Information, Communication & Society*, *23*(1), 128–147. https://doi.org/10.1080/1369118X.2018.1486870
[2] Robinson, E. P., & Zhu, Y. (2020). Beyond "I Agree": Users' Understanding of Web Site Terms of Service. *Social Media + Society*, *6*(1), 1–13. https://doi.org/10.1177/2056305119897321

reporting harmful content from users, complaint systems for content removals; ▶chapter 4) (TCO Regulation, Art. 7.1).

Use designation lists and consider clearly banning other categories of harmful content forms

A variety of content types can be harmful to users, to society and to the platform itself. Thus, HSPs should consider the following additional prohibitions, also addressed by other legal frameworks like the Digital Services Act, and, thus, prohibit content containing for example hate speech or incitement to violence. At the actor level, terrorist groups may be excluded from platform use. Designation lists (i.e. official lists of terrorist groups published by democratic states) are a good resource for such bans. Furthermore, competent authorities may alert HSPs to content which is not classifiable as terrorist but nonetheless harmful; assessing the content against the ToS and the eventual handling of non-terrorist but harmful content is ultimately the responsibility of HSPs (TCO Regulation, 40).

Communicate accepted use

ToS should also describe, by reference to platforms' values and purposes, what usage behaviours are not simply accepted but also welcomed and encouraged by HSPs'. There is no universal standard for this – it varies greatly depending on the platform's functionalities and principles. Having said this, describing accepted use is not explicitly required to meet the TCO Regulation's obligations.

Provide information on how to report prohibited content

One way of surfacing illegal and ToS violative content is through user reporting systems. HSPs should therefore include a section in their ToS explaining processes for reporting content suspected of violating ToS.

Determine the consequences of a ToS violation

To ensure ToS violations are dealt with transparently and to remain accountable to the public, HSPs should clearly set out in their ToS the actions to be taken against users and/or content for ToS violations. Per the TCO however, a removal order from a competent authority must result in content removal within one hour of its being flagged. Removal orders may also be actioned by blocking the content or geo-blocking it in the EU.

Publish annual transparency reports

When preparing and revising ToS, HSPs should publicly commit themselves to regularly publishing transparency reports. Practical tips and TCO regulatory requirements for creating transparency reports follow in ▶chapter 6.

Chapter 2

Specific Measures for Identifying and Removing (Terrorist) Content

Summary: Contents and Main Points of this Chapter

- The ability of HSPs to identify content as terrorist is fundamental to implementing the TCO Regulation and beyond.
- In many cases, the decisions to be taken are not easy and can be **very contextual and sensitive.**
- F**undamental rights such as freedom of expression must be carefully taken into consideration**.
- **When it comes to the TCO Regulation, HSPs do not have to assess the legality of content alerted by a competent authority**. However, once a removal order has been issued, both **HSPs and content providers (i.e. users) have the right to challenge it**.
- It is advisable to set up a process that is followed when identifying terrorist content. Such a process will comprise multiple steps.
- Various practical considerations, such as the use of designation lists or symbol databases and methods for conducting the requisite balancing exercises, are involved in the assessment of content.
- Furthermore, review processes play a particularly important role in the special case of **cross-border removal orders**, where a HSP does not receive the removal order from the competent authority in which it has its main establishment or legal representative. **Special procedures apply** in such cases.

It is of fundamental importance that HSPs are able to assess whether content is illegal or terrorist in order to detect terrorist content proactively, as well as to legally challenge removal orders when HSPs believe they may have been made erroneously. Correctly identifying content as illegal and terrorist in nature is important in order to preserve otherwise legal and non-terrorist content online as well. Such decisions are often challenging, and it is **always necessary to weigh the harm of content against the harm of breaching the fundamental right to freedom of expression**.

1. Establishing identification processes for illegal and harmful content

Identification processes comprise a defined series of steps companies or platforms take to determine whether content violates the TCO Regulation or elements of platforms' ToS. These steps should guide a methodical assessment and enable a reasoned conclusion on the illegality of content.

Such balancing decisions are **necessary to** (a) **take proactive measures** against terrorist and/or otherwise harmful content and (b) **challenge contentious cases of removal orders** issued by the competent authority on the basis of the TCO Regulation (more information on the process of legal remedy will follow in this chapter in ▶subchapter 4).

Proactive measures against terrorist and other harmful content are recommended for various reasons.

1. By implementing such measures HSPs demonstrate reliability and goodwill towards political decision-makers, competent authorities, and other relevant stakeholders.
2. Proactive measures build trust with stakeholders because they positively de-limit the legal liability or reputational damage that might be incurred by terrorist exploitation of platforms.
3. Insufficient or inadequate action against terrorist, extremist or otherwise harmful content may result in the platform attracting users who intend to distribute exactly such content. The consequence of this would be an increased prevalence of prohibited content that, in turn, requires greater investments of time, effort and financial resources on the part of HSPs to defeat it.

2. A process for identifying terrorist content

1) Definition of terrorist content

Any classification of content is undertaken by reference to a definition. As far as the TCO Regulation is concerned, the definition of terrorist content is that contained in EU Directive 2017/541 as mentioned in the ▶introduction. It defines terrorist content as textual, visual, or auditory

> "material that incites or solicits someone to commit, or to contribute to the commission of, terrorist offences, solicits someone to participate in activities of a terrorist group, or glorifies terrorist activities including by disseminating material depicting a terrorist attack. The definition should also include material that provides instruction on the making or use of explosives, firearms or other weapons or noxious or hazardous substances, as well as chemical, biological, radiological and nuclear (CBRN) substances, or on other specific methods or techniques, including the selection of targets, for the purpose of committing or contributing to the commission of terrorist offences." (TCO Regulation, 11)

The definition does not include "material disseminated for educational, journalistic, artistic or research purposes or for awareness-raising purposes against terrorist activity" (TCO Regulation, 12), and fundamental rights such as freedom of expression, information and science shall always be carefully taken into consideration. "Furthermore, the expression of radical, polemic or controversial views in the public debate on sensitive political questions should not be considered to be terrorist content." (TCO Regulation, 12)

2) Support through the use of automated tools

It can often be helpful to use automated tools for preliminary identification of potentially problematic content. For example, automation can be used to identify specific keywords associated with

extremist worldviews or visual elements such as logos or symbols of terrorist organisations. There are also initiatives to automatically detect potentially terrorist content across platforms by reference to a common database (e.g. Tech Against Terrorism's Terrorist Content Analytics Platform [TCAP] or the hash-sharing database maintained by the Global Internet Forum to Counter Terrorism [GIFCT]).

3) Support from reporting systems

If a HSP is exposed to terrorist content – formally this means if the competent authority of the country hosting the HSP's main establishment or the legal representative has issued at least two legally binding removal orders and the HSP has been informed of this – the TCO Regulation explicitly requires measures to be taken to prevent the further dissemination of such content (TCO Regulation, Art. 5.2). Preventive measures are also recommended to demonstrate proactive behaviour. One such measure can be a reporting system that allows users to report suspicious and prohibited content to the HSP. It makes sense to have users specify a category they would assign to suspicious content with terrorism as one such category. This allows employees who process the reports to prioritise and process content suspected of spreading terrorist ideas more quickly. This pre-categorisation can also facilitate the preparation of later transparency reports – more on this in ▶chapter 6.

4) Assignment of and review by human moderators

Flagging content through automated tools or reporting by individual users is the first step in drawing attention to suspicious, potentially dangerous content. Subsequently, however, review by human moderators is usually necessary until a final decision on the action to be taken is reached. These moderators should (1) be well-versed in the regulations applicable to the platform, (2) be able to differentiate prohibited content from material that engages the right to freedom of expression, and (3) have precise knowledge of different moderation options on the platform. It is important that human moderators are constantly trained on terrorists' online strategies, but no less important from an ethical perspective that they are counselled in the psychological effects of dealing with problematic content.

5) Decision on how to deal with the content

The process outlined above culminates in a decision on how to deal with the content. It may be that nothing is done, and the content remains active if, once all the considerations have been weighed, the content is deemed to be harmless. At the other extreme, in the case of highly problematic, dangerous or otherwise prohibited content, such as calls for terrorist attacks, content and users may be blocked. In between, however, there are also numerous other ways to proceed with content – content that is not in scope of the TCO Regulation may require a more nuanced treatment. Details and practical suggestions can be found in the following ▶third chapter on effective moderation strategies.

6) Notification of the content provider

If content is blocked, the content provider must be notified. An automatic notification and complaint system is recommended for this. Generally, a process suitable for the HSP should be developed and set up for this purpose. Details and suggestions for how this might be done in compliance with the TCO Regulation can be found in the ▶ fifth chapter.

3. Practical tips and advice: What helps to assess whether the content is illegal?

Regarding the TCO Regulation, HSPs do not have to assess the legality of content as it is the responsibility of the competent authorities to do so before issuing a removal order. However, **HSPs are required to proactively identify terrorist content on their platforms once they have been exposed to terrorist content**. It is often challenging to assess whether content crosses the legal line and should therefore be prohibited. In addition, the psychological stress caused by seeing problematic content should not be neglected. Employees who moderate content should have regular opportunities to reflect on their work and, if needed, receive psychological support.

To assess the illegality of content, you will find some clues and practical tips below. Before that, let's take a look at the relevant parts about content assessment in the TCO Regulation:

Paragraph 11 states, among other things:

> "When assessing whether material constitutes terrorist content within the meaning of this Regulation, competent authorities and hosting service providers should take into account factors such as the nature and wording of statements, the context in which the statements were made and their potential to lead to harmful consequences in respect of the security and safety of persons." (TCO Regulation, 11)

In addition, fundamental rights must always be weighed up:

> "When determining whether the material provided by a content provider constitutes 'terrorist content' as defined in this Regulation, account should be taken, in particular, of the right to freedom of expression and information, including the freedom and pluralism of the media, and the freedom of the arts and sciences. Especially in cases where the content provider holds editorial responsibility, any decision as to the removal of the disseminated material should take into account the journalistic standards established by press or media regulation in accordance with Union law, including the Charter." (TCO Regulation, 12)

If HSPs receive a removal order, the content has already been classified as terrorist by the competent authority. In **two scenarios it is useful** for HSPs to be competent to **be able to classify content** as terrorist or non-terrorist:

1. If a HSP has not yet received an official removal order but wants to **act proactively** and/or enforce compliance with its own ToS in respect of suspicious content.

2. If a HSP receives an official removal order but **has doubts** about the assessment of the content as terrorist by the competent authority and wants to review the order (as a first step towards a legal remedy).

Below you will find various forms of practical assistance to classify content as terrorist or non-terrorist.

Use designation lists

National and international designation lists that name terrorist organisations provide a good framework and reference point for content classification. For example, adopting designation lists as the basis for banning or blocking, which should be referenced in the ToS (▶chapter 1), gives a platform's sanctions the backing of the law. Paragraph 11 of the TCO Regulation explicitly suggests that the European Union list may be used when assessing content.

Make use of keyword, symbol and logo databases

People who manually review the content should not only have the definition of terrorist content readily available but also be conversant with and maintain their knowledge of keywords and phrases typically used by terrorist organisations. The more intensively a person assessing the content is familiarised with a terrorist phenomenon (e.g., right-wing, left-wing, Islamist), the easier it is to recognise obfuscation tactics such as so-called 'dog whistling', i.e. the use of seemingly harmless words that have ideological meanings within the scene[3]. A database of logos and symbols, i.e. visual elements that indicate a terrorist background, is equally indispensable for assessing content accurately and at pace.

Include contextual factors

Moderation should consider the context in which content is likely to have been published. Relevant contextual factors include but are not limited to (a) political conditions, (b) current events, including recent developments in the news, and (c) cultural circumstances that may shape views on certain issues. This is often difficult to judge – especially when users are anonymous, and the content they post contains few or no textual clues to their identities – but in some cases, when external factors are referenced in content, context considerations turn out to be helpful.

Additionally, when dealing with **content that is not terrorist but harmful in another way** (e.g. different forms of hate speech), the following points may be considered.

Weigh the extent of potential damage caused by the content

Terrorist, extremist and further harmful content like hate speech or incitement to violence, especially when it incites the commission of terrorist offences, can cause significant harm. When assessing content, taking into account the extent to which content may be a causative factor in such damage can be useful. The greater the damage that may occur, the faster and

[3] Åkerlund, M. (2022). Dog whistling far-right code words: the case of 'culture enricher' on the Swedish web. Information, Communication & Society, 25(12), 1808–1825. https://doi.org/10.1080/1369118X.2021.1889639

more deliberate the action that must be taken. Adverse psychological effects on those affected by the harmful content should be considered a form of damage.

Consider the potential impact of content removal

The TCO Regulation calls for a sensitive balancing of the fundamental rights enshrined in the Charter (e.g. freedom of expression and information) when considering content removals. If removal has not been ordered by a competent authority, there are other ways in which harmful content can be handled – examples of which can be found in ▶chapter 3.

4. What should I do if I see it differently? How to challenge a received removal order

What should be done when the assessments of the HSP and the competent authority conflict?

HSPs as well as the content provider have the right to challenge removal orders under Article 9 of the TCO Regulation. This right is an important check and balance capable of upholding fundamental rights.

Challenges to removal orders are carried out in the courts of the EU Member State whose competent authority ordered the removal. **HSPs are obliged to keep in secure conditions and for a period of six months all content that has been removed (either as a result of a removal order or other measures), as well as any data associated with the content (e.g., time of publication, account information).** This aids the investigation and prevention of terrorist offences or threats and ensures that there is sufficient time to initiate legal challenges to the content removal and, if necessary, to restore the content (TCO Regulation, 27 & 28).

Special case: Cross-border removal orders (TCO Regulation, Art. 4)

If HSPs receive a removal order from a competent authority in an EU Member State other than its 'home authority' (i.e. the competent authority of the Member State in which the HSP has its main establishment or legal representative; details on legal representatives follow in ▶chapter 4), **a special procedure applies.**

When issuing the removal order, the competent authority must also send a copy of that order to the 'home authority', which has a period of 72 hours to review the order. If the removal order is found to infringe on fundamental rights and freedoms enshrined in the Charter, it may issue a reasoned decision, which may constitute an objection.

The HSP must remove and secure the content upon receipt of the removal order (as in the case of a removal order by the 'home authority'). The HSP (and the provider of the removed content) may submit a request for review to the 'home authority' within 48 hours, which is entitled to review the removal order and respond within 72 hours of receiving the application, having notified the competent authority that originally ordered the removal that a review is being undertaken.

After the reviewing authority has issued a reasoned decision, the relevant actors (namely the original authority, HSP, content provider, and Europol, if applicable) are notified and, if a breach has been detected, the HSP can immediately restore the content.

Chapter 3

Establishing Effective Moderation Mechanisms for Terrorist Content Online

Summary: Contents and Main Points of this Chapter

- In content moderation, **user content is reviewed to assess whether legal** (e.g. TCO Regulation, DSA) **and platform-specific regulations** (e.g. ToS) **have been complied with or violated**.
- **If a rule has been violated, content will be 'moderated'.** In other words, measures are taken to limit the reach of the content.
- When receiving a **removal order** from a competent authority based on the TCO Regulation, **content moderation always includes the removal** of the respective terrorist content, though HSPs and users can contest this if they disagree.
- Various points must be taken into account before, during or shortly after the moderation measures are undertaken by the HSP, as well as in a later review. These include the notification of users whose content has been moderated and the option to appeal the decision.
- If platforms take **proactive action against other potentially harmful forms of content**, such as hate speech or insults, regardless of the TCO Regulation, other **alternative moderation approaches** may also be considered.

1. What is content moderation and why is it necessary in some cases?

Content moderation involves **reviewing user-generated content** (UGC) on the internet in order to **assess the appropriateness** of the content based on platform regulations (e.g. ToS) and legal frameworks (e.g. TCO Regulation, the DSA)[4]. **Platform regulations and legal requirements are enforced through content moderation** so that the requisite action is taken against prohibited and/or problematic content. HSPs are required to **be transparent** about specific measures taken to identify and remove terrorist content, including proactive and reactive content moderation processes and any automated tools used.

Content moderation is a very important, but also very complicated and **sensitive** procedure. Having the appropriate technical and subject matter resources can be a significant financial burden for smaller companies. In addition, terrorist content can be produced in all languages and it is extremely resource-intensive to ensure that moderators are able to assess content in all these languages.

[4] Roberts, S.T. (2017). Content Moderation. In L. Schintler & C. McNeely (eds.): *Encyclopedia of Big Data*. Springer, Cham. https://doi.org/10.1007/978-3-319-32001-4_44-1

The fundamental rights enshrined in the UN Declaration on Human Rights, in particular the right to freedom of expression, must always be weighed against the harm caused by the content in question. On the other hand, HSPs have a social responsibility to mitigate harmful content on their platforms.

Comprehensible and transparent content moderation helps to defend users and stakeholders against the effects of harmful content and builds trust. From the perspective of HSPs, content moderation can bolster limitations of liability, compliance with applicable laws and protection against reputational damage caused by misuse of the platform.

Content moderation does not necessarily mean content removal. While content deletion can be mandatory due to legal requirements – as is the case with removal orders based on the TCO Regulation – there are other ways to deal with harmful content. Depending on how problematic and serious the content is and the damage it may cause, alternative options can be considered. Examples and suggestions of alternative moderation strategies can be found in ▶section 3 of this chapter.

Moderation styles vary from platform to platform depending on their functionality, services offered, platform values and risk tolerance[5]. It is important to be as transparent as possible with content moderation. This includes various factors that need to be considered before, during and after moderation takes place to ensure users know the moderation actions their content might incur.

1. **Before:** 'Warn' the users in advance

Let users know at an early stage what content moderation actions might occur following conduct or content that is prohibited by ToS or by law. This means: Include information about your moderation strategies in the binding agreement with the users, i.e. in the ToS (▶chapter 1).This also helps to ensure that you begin to build trust with individuals before or at the time of entering into the relationship with the user and that you give a binding and unambiguous assurance that problematic and illegal content will not be tolerated on the platform and that you, as a HSP, want to protect users from such content.

[5] Roberts, S.T. (2017). Content Moderation. In L. Schintler & C. McNeely (eds.): *Encyclopedia of Big Data*. Springer, Cham. https://doi.org/10.1007/978-3-319-32001-4_44-1

Funded by the European Union

2. **During (or shortly after the action):** Notify users whose content has been blocked or moderated and offer the option to appeal the decision

Article 11 of the TCO Regulation obliges HSPs firstly, to inform the user who provided (i.e. authored and/or uploaded) the prohibited (terrorist) content that it has been blocked by the platform, and, secondly, to provide information to the user about the background or the removal order at the user's request. Notification can be withheld from content providers temporarily for a period of no longer than six weeks, if the competent authority considers that there is a particular danger involved in communicating the removal to the content provider. When non-disclosure still remains important and appropriate, the competent authority may extend the period by further six weeks (TCO Regulation, Art. 11.3). The TCO Regulation (Art. 10) provides specific mechanisms for handling such situations. It is advisable to set up a standardised automatic process by which content providers are notified of deletions and are given an opportunity to file a complaint against them. This notification and objection mechanism should also be deployed in cases that do not fall under the TCO Regulation, but have taken place through (pro-)active measures on the part of HSPs.

3. **Afterwards:** Regularly communicate key data on HSP-wide content moderation measures publicly

Transparency about content moderation decisions and outcomes is important to establish trust and accountability between HSPs and their users, and is increasingly mandated by different types of online regulations. Transparency reports capture information on platforms' moderation decisions, including not just the number and type of violations identified but also exactly how such violations were handled. (More details on transparency reports and TCO transparency requirements can be found in ► chapter 6.)

2. Practical tips and advice: How should content moderation be implemented?

It is important for HSPs to have a clear **process for content moderation** that can be **customised** to specific business services and needs. It is crucial to first identify prohibited content and to be able to assess its illegality, as discussed in more detail in ► chapter 2. As a reminder, these are the key steps:

1. Definition of terrorist content
2. Support through the use of automated tools
3. Support from reporting systems
4. Assignment of and review by human moderators
5. **Decision on how to deal with the content (i.e., decision about content moderation)**
6. Notification of the content provider

In this section we are concerned with *step 5*. In ► chapter 2, we named this step 'Decision on how to deal with the content'. Having received a removal order from a competent authority, HSPs must

remove the content and have the possibility to appeal the order. HSPs can also decide to identify and remove terrorist or prohibited content proactively (i.e. absent removal orders).

This guide distinguishes five different types of moderation: pre-, post-, reactive, distributed and automated. Brief explanations of each, as well as their advantages and disadvantages can be found below.

Type of moderation	Explanation
Pre-moderation	Moderators review the content before it is published. **Pros:** High degree of safety and content compliance with legal standards and platform-specific regulations **Cons:** High burden of effort, personnel and (eventual) financial costs
Post-moderation	Moderators review the content immediately after publication. **Pros:** Enabling quick user interactions with the content (due to immediate publication) **Cons:** Significant amount of time, personnel (and eventual) financial costs; in the case of prohibited/harmful content, users are exposed to the content
Reactive moderation	Moderators review the content after users have reported it. **Pros:** Less resource-consuming HSP effort; Building trust with users (through the possibility of reporting) **Cons:** Responsibility of the users; in the case of prohibited/harmful content, users are still exposed to the content; potential for false alarms requiring additional effort on the part of the HSP
Distributed moderation	The user community acts as moderators, often through an up- and down-vote mechanism which measures the trustworthiness of the content and is ultimately used to determine the content's public reach (i.e. content with many up-votes ranks better and gets a higher reach). **Pros:** Less resource-consuming on the part of the HSP; Encourage user engagement; Self-regulation **Cons:** Responsibility of the users; in the case of prohibited/harmful content, users are exposed to the content; Susceptibility to coordinated manipulative behavior on the part of malign users
Automated moderation	Models based on artificial intelligence (e.g. filters, algorithms) function as moderators. **Pros:** Less resource-intensive once it is set-up; early, fast, highly scalable detection of potentially malicious content

	Cons: Susceptibility to errors, especially with regards to erroneous content deletion (problematic in the context of freedom of expression, thus human assessment important); continuous maintenance and adaptation to new developments necessary

(Based on Grimes-Viort, 2010)[6]

It is recommended to combine several approaches to content moderation. Reactive moderation, for example, can be easily combined with pre- or post-moderation. In most cases, **combinations are very useful, and in some cases can even be necessary**. Automated moderation decisions must always be reviewed or at least informed by human review to avoid systematic breaches of users' fundamental rights and risks to freedom of expression.

Be aware that extremist and terrorist actors can use **tactics to avoid known content moderation methods**, operate under the radar, and thus circumvent moderation, particularly automated content moderation mechanisms. Popular tactics include the use of *URL shortening* to evade filters or websites blocks, account and content *mirroring* which involves posting or creating identical content/accounts multiple times to overwhelm moderators and have back-ups available in case of deletion, or *deliberately choosing (incorrect) spellings of a word* to slip past automated word filters. The range of evasive techniques emphasise the importance of human moderation.

Your involvement with this guide, or your (certified) participation in the online course which is also offered as part of the TATE project, show that you are interested in these trends. You can find more examples of content moderation evasion as well as feasible and effective responses in Tech Against Terrorism's Knowledge Sharing Platform, which is available here.

3. Alternative moderation approaches

Moderation doesn't always have to mean content deletion. Alternative approaches, of the kind suggested below, may be of interest to HSPs seeking to moderate content (pro-)actively beyond the requirements of the TCO Regulation. It's important to highlight that such alternative moderation approaches are **out of scope of the TCO regulation and can only be used when platforms have not received a removal order but want to moderate non-terrorist, otherwise harmful content proactively.** When you receive a removal order, the procedure is clear: You have to remove the content and no alternative moderation approaches come into question.

[6] Grimes-Viort, B. (2010, December 7). 6 types of content moderation you need to know about. *Social Media Today*. https://www.socialmediatoday.com/content/6-types-content-moderation-you-need-know-about

WANT TO DO MORE TO KEEP YOUR PLATFORM SAFE?

Hide content

HSPs can partially or completely hide content, and thereby avoid blocking it, if they believe that users may find the content offensive or objectionable but it is nonetheless legitimate, legal and permissible across the entire platform. Hiding content from people from a vulnerable group or people located in a country where the content is illegal (while it is allowed in other countries) is one such response. Various technical functionalities can be used to hide content, such as login or paywall filters, a secure (search) mode to display age-appropriate content and geo- or time-blocking.

Decouple content from reward mechanisms ('disengagement')

Disengagement deprives certain content or users of engagement metrics, deters activity around the post, and can make content generally unrewarding to post. However, the content and user account remain on the platform. Disengagement restricts the prominence of posts or accounts on the platform. Typical disengagement tactics include the disabling of platform features, such as, in the case of many social networks, the ability to like, comment, or share posts, so that the post can only be read, demonetisation (i.e. depriving accounts of the ability to make money from their content), or de-verification (i.e. removing any certification of the account's or user's identity). Such sanctions can entail, and be compounded by, a change in the treatment of the content or account by the platform's algorithm: downgrading the content means that it is more difficult to be widely distributed and promoted through the platform's mechanisms (usually recommendation algorithms).

Attach pedagogical notes to the content

The goal of pedagogical or communication-based tactics is to offer users additional information so that they can ultimately decide for themselves whether they want to see the content or not. In the end, the platform decides which content is provided with such notes, what these notes involve, what category of harm users are warned about, and how much additional information is offered. A well-known practice previously used by (previous) Twitter is to alert users that there may be harmful content in a post, such as misinformation or conspiracy narratives, and to permit users to see the content only after they have actively confirmed that they wish to do so by clicking on a button. Particularly in the case of political-ideological content that might promote radicalisation, counter-narratives and links to educational information can sensitise people to the possible effects of such content.

Give responsibility to users ('community empowerment')

The premise of moderation mechanisms based on community empowerment is to allow the users themselves to create the digital space they imagine. Such strategies

> may be of particular interest to platforms where the idea of community is important, or whose moderation practices are already to an extent dependent on the support of users. These moderation approaches follow the type of distributed moderation. In addition to the up and down voting functionality already mentioned, this also includes the individual blocking or muting of specific accounts, which a large number of platforms already offer, or the use of admins or moderators from the community itself. Closely linked to this is the concept of 'Trusted Flaggers', which is examined in the Digital Services Act (Article 22). The concept refers to users who are particularly trustworthy and competent to assess the illegality of content and report it (objectively and quickly) and who represent collective (public welfare-oriented) interests regardless of the specific online platform. Content reported in this way should be processed on the HSP side as a matter of priority and quickly.

These alternative approaches to moderation can be relevant even where platforms are not compelled by the TCO Regulation to take action. Regardless of the form it takes, the TCO Regulation accommodates a proactive approach: if, in the course of (pro-)active, own moderation measures, the HSP encounters content that deals with an imminent threat to life or a terrorist act, this must be deleted and the competent authority of the EU Member State affected by it must be informed immediately (TCO Regulation, Art. 14.5).

More details on the (technical) methods required by these alternative approaches, as well as their advantages and disadvantages and case studies, are provided by Tech Against Terrorism here.

Chapter 4

Establishing Points of Contact and Legal Representatives

Summary: Contents and Main Points of this Chapter

- The TCO Regulation distinguishes between contact points and legal representatives.
- **Contact points must be set up by each HSP** and are **responsible for** receiving **removal orders** and processing them promptly.
- **If HSPs are not established in the EU, a legal representative must also be appointed.** This person is responsible for receiving, complying with and enforcing the TCO Regulation. The legal representative may, but does not have to, act as a contact point simultaneously.
- Every HSP, whether or not it has been exposed to terrorist content, must establish a point of contact and, where necessary, a legal representative as per TCO Regulation.

The TCO Regulation stipulates that *all* HSPs that are affected by this regulation (i.e. fall within the scope of the regulation by definition; ►see the section on affected platforms in the introduction) are obliged to appoint a contact point or legal representative.

1. What are contact points and legal representatives?

Contact Point (TCO Regulation, 42 & Art. 15)

- **Purpose**: HSP contact points facilitate the immediate processing of removal orders. The contact point therefore serves only operational purposes.
- **Logistics**: The contact point should be able to receive and transmit removal orders electronically, whether it is located in-house or outsourced.
- **Necessary resources**: The contact point must have sufficient technical ability, access, and be staffed in such a way that removal orders can be processed without delay. Since terrorist content must be removed within one hour after receiving a removal order, this means that the contact point must be available 24/7.
- **Location**: The contact point does not necessarily have to be located within the EU.
- **Communication**: The language in which to communicate with the contact point should be indicated in the information about its availability. In order to enable communication between HSPs and Member State competent authorities, at least one official language of the EU should be used. This should be a language in which the ToS of the platform are also available.

Legal representative (TCO Regulation, Art. 17)

- **Necessity**: If the HSP does not have its main establishment in the EU, a legal representative must be designated. This is a natural or legal person who is located in one of the EU Member States in which the HSP offers its services.
- **Purpose**: This legal representative is responsible for receiving, complying with, and enforcing the TCO Regulation and removal orders in particular.
- **Resources**: HSPs must provide legal representatives with the power, capacities, and resources to comply with the TCO Regulation and to cooperate with the competent authorities.
- **Liability**: The legal representatives can be held liable for violations of the TCO Regulation.
- **Relationship with the contact point**: The legal representative can also act as a contact point at the same time but is not required to do so.

2. Why is it necessary to have a contact point or legal representative?

The other chapters discuss many options that may be appealing to HSPs from a business standpoint either, but the main reason and argument to **set up a contact point or designate a legal representative** is because it **is mandatory**. This may involve some reallocation of resources, but it is unquestionably in HSPs' best interest to comply with the law and avoid the negative consequences of non-compliance, such as loss of reputation, trust, and money, if fines are imposed, by making the appointment promptly.

HSPs must allow competent authorities to review the information about the contact point and subsequently to provide electronic notifications (i.e. to submit a removal order electronically). Typically, this involves providing an e-mail address (for example, in the "Contact Us" section of the HSP's website) through which the competent authority can approach the contact point.

There are two aspects of appointing a legal representative which must be emphasised. Firstly, the **legal representative's identity must be publicised** (see contact point). Secondly, **the HSP must actively notify the appointment to its 'home authority'** (i.e. the competent authority of the Member State where the legal representative is established).

Europol has also developed a platform, Plateforme Européenne de Retraits de Contenus illicites sur Internet (European platform for takedown of illicit content online), or **PERCI, to support the implementation of the TCO Regulation**. The purpose of PERCI is to ensure HSPs can receive removal orders from Member States through a common secure channel, instead of 27 separate systems per Member State. It streamlines removal orders from different Member States and acts as a single point of contact in the context of referrals and removal orders and is intended to prevent the duplication of removal orders, i.e. the same order sent by two Member States. It also is useful for HSPs to have a centralised receipt of removal orders and referrals received over time, to support transparency reporting obligations. Through PERCI, HSPs are also able to request scrutiny and review to challenge a removal order.

3. What is the competent authority of and EU Member State and how do I contact them?

Contacting the competent authority about the legal representative is necessary in certain circumstances. One such 'circumstance' arises when challenging a removal order. Another case would be, for example, if the HSP becomes aware of terrorist content involving an imminent threat to life or a terrorist act without a removal order. In this case, the HSP is required to delete it immediately and inform the competent authority of the EU Member State affected by it (TCO Regulation, Art. 14.5).

Most competent authorities of EU Member States have already set up contact points. An up-to-date list including contact details can be found on the website of the European Commission here.

Chapter 5

Setting Up a User Notification and Complaint System for Removed Content

Summary: Contents and Main Points of this Chapter

- Affected **users can use complaint procedures to appeal removal orders** (and, if necessary and beyond the scope of the TCO, other proactive moderation measures).
- Complaint procedures are important as a **control and feedback mechanism**, from the perspectives of users, companies, and the law.
- HSPs "shall establish effective and accessible" complaint mechanisms (TCO Regulation, Art. 10.1).
- According to the TCO Regulation, complaint systems must meet certain content and technical requirements.
- A complaint procedure can result in two different outcomes, namely (1) the complaint is upheld because the content is found to have been erroneously blocked, or (2) the complaint is dismissed because the content is found to have been blocked justifiably.
- Depending on the outcome of the complaint process, the content may be subject to other measures.
- This chapter additionally provides guidance on designing and implementing a HSP-specific complaint procedure.

Complaint procedures allow users to appeal content removals by communicating with the platform and **are the first step towards (legally) challenging a removal order**. You can read more about the dispute process in ▶ chapter 2.

1. Why is it necessary to set up a transparent complaint mechanism?

It is important to set up a complaint mechanism, specifically from (a) a legal perspective, (b) a user perspective, and (c) a company perspective.

a) Legal perspective

Complaint procedures conform to **regulations established in law** which **make such mechanisms necessary**. For example, the TCO Regulation (TCO Regulation, Art. 10) stipulates that HSPs must establish an effective and accessible complaint mechanism to give users the opportunity to contest the removal or blocking of the content after a specific measure. Under the TCO Regulation, the content provider must be informed of the final outcome within two weeks.

b) User perspective

A complaint mechanism is not simply a legal requirement. A clear and accessible complaint mechanism helps build trust with users – both with those whose content has been moderated and, in particular, with unaffected users who use the platform as intended. In this way, HSPs demonstrate that users can rely on moderation processes, that such processes are founded in a sense of **responsibility towards users**, and that platforms are considerate of **fundamental rights** such as freedom of expression and information.

c) Company perspective

Complaint mechanisms can be a helpful form of **self-monitoring**, whereby moderation measures and standards can be assessed for efficacy, fairness, and consistency. Such mechanisms can provide an assurance that your platform is used in the way it is intended to be, which further helps to protect the **reputation** of online services, and further protect the right to freedom of expression online.

2. What are the requirements for a complaint system?

According to the TCO Regulation (33 & Art. 10), complaint systems should:

- Be user-friendly,
- Be effective and (easily) accessible,
- Provide a secure framework in which complaints are dealt with promptly and transparently, so that the complainant is informed of the outcome of the review within two weeks.

The complaint systems must be set up for the purpose of the TCO Regulation to restore erroneously removed content. However, complaints can also be heard against measures taken to enforce the platform's ToS beyond the scope of the TCO Regulation.

3. How are complaints to be handled and what are the possible results?

Once the complaint system is implemented and users have submitted a complaint, the HSP will **review the complaint** and **communicate the decision to the complainant within no more than two weeks for cases engaging the TCO Regulation**.

There are two possible outcomes of the review process, listed in the following table.

Outcome A **User complaint against content removal upheld**	*Outcome B* **User complaint against content removal dismissed**
Result: The complaint against the content removal is **justified** and it is therefore allowed.	**Result:** The complaint against the content removal is **not justified** and is therefore dismissed.
Meaning: The content has been improperly removed, deleted or otherwise moderated.	**Meaning:** The content has been properly removed, deleted or otherwise moderated.
Further procedure: The HSP (1) informs the complainant of the result of the review and (2) restores the content.	**Further procedure:** The HSP (1) informs the complainant of the result of the review and (2) gives the user the reasons for this decision.

4. Practical tips and advice: What elements are useful when establishing a complaint system?

That a complaint system must have certain features is a requirement not just of the law, but of **user-friendliness**.

Like other key issues, such as ToS (▶chapter 1), the establishment of a process for identifying prohibited content (▶chapter 2), or the choice of platform-specific moderation mechanisms (▶chapter 3), the method of submitting a complaint can vary greatly between platforms. While some platforms may offer e-mail complaints, others choose to implement standardised online form-based applications. **Complaint systems should be tailored to a platform**'s purpose, structure, and organisation. Guidance on setting up a complaints system can be found in the following **checklist**.

Give clear information about the content removal

Notify the person whose content has been removed. Consider doing this even if the content has been moderated in some other way. In this context, also inform the person about why the content was removed (see below the following point on education and pedagogical background information) and how an appeal can be lodged against the removal decision (see the following point on explaining of the complaint process).

Explain the complaint process

Explain the complaint process to your users. This should be done in the course of giving notification that content has been removed or otherwise moderated. Information about the complaint process can also be included in the ToS. The explanation of the complaint process should include: (1) how complaints can be submitted, (2) how the review process works, and (3) how users are notified of the outcome of the review.

Provide pedagogical background information

Provide users with educational opportunities to explain why the content was removed and which terms it violates. For example, the content may be removed if it violates the ToS or if a removal request has been received from a competent authority as part of the TCO Regulation. In the former case, this information can be specified by adding which of the ToS platform-specific prohibitions has been violated (e.g. against hate speech, incitement to violence, sexual content and harassment). In the latter case, it is recommended to add a brief informative overview of the TCO Regulation to the complaint system, so that the legal framework is made clear to the user.

Give regular updates on the progress of the complaint

Users should be provided with regular updates on the progress of the complaint to show that the process is continuing. At a minimum, such updates should consist of a notification when the review has been completed and timely communication of the outcome. More detailed communication might notify the user that the complaint has been received and is now being reviewed by HSP employees. Best practice would give the user a timeline for the adjudication of their complaint. These updates can be given by e-mail or displayed in an online portal.

Document the (individual) complaint process

It is important to document the complaint process. This document will serve as a reference for any subsequent questions or disputes.

Chapter 6

Practical Support and Advice Around Transparency Reporting

Summary: Contents and Main Points of this Chapter

- Transparency reports enable HSPs to **publicly communicate** how their **values** are being upheld on the platform, as well as **actions taken against prohibited and illegal content and conduct**.
- Transparency reports are an **important tool for taking public responsibility and demonstrating credibility and trustworthiness**.
- Different **laws**, including the TCO Regulation, explicitly **require transparency reporting**.
- When preparing transparency reports, it is advisable to proceed systematically before, during and after the process by following specific steps.
- The **TCO Regulation requires HSPs that are exposed to or have taken action against terrorist content to publish an annual transparency report on their activities in relation to terrorist content**. The report must include certain essential figures and information, such as the number of pieces of content removed. The report must be published by March 1 of the following year at the latest (TCO Regulation, Art. 7.2).

1. What are transparency reports?

Transparency reports are an **important tool for HSPs to prove their accountability, credibility, and trustworthiness, and to publish socially relevant information**. Transparency reports **contain important data about requests that HSPs receive from state actors worldwide and how these requests are handled**, which is intended to make collaboration and cooperation with authorities and other state bodies transparent[7].

Transparency reports also provide an overview of what measures the HSP has taken to enforce regulations (e.g. through content removal or other moderation measures). This includes the enforcement of (1) platform-specific policies (usually ToS; ▶ chapter 1), (2) rights such as copyright or trademark law, and (3) (local) legislation and regulations that result in content removal[8]. In the

[7] Urman, A., & Makhortykh, M. (2023). How transparent are transparency reports? Comparative analysis of transparency reporting across online platforms. *Telecommunications Policy, 47*(3), 102477. https://doi.org/10.1016/j.telpol.2022.102477

[8] Trust & Safety Professional Association (2023). What Is A Transparency Report? *TSPA*. https://www.tspa.org/curriculum/ts-fundamentals/transparency-report/what-is-a-transparency-report/

EU, local legislation includes the TCO Regulation and the Digital Services Act. Country-specific legislation may also be relevant, such as the Network Enforcement Act in Germany.

Transparency reports are typically published on a regular basis. The TCO Regulation states that this should be done **(at least) once a year** if a HSP has taken measures against terrorist content (TCO Regulation, 30).

From HSP to HSP, transparency reports and the metrics reported in them can vary greatly[9]. **For transparency reports in line with the TCO Regulation, there are specific requirements about what information needs to be included.** You can read more about this in ▶ section 4 of this chapter.

2. Why are transparency reports necessary?

Transparency reports **allow users and third parties to assess the extent to which HSPs remain true to their own principles, legal requirements, and data protection and privacy**[9]. Let's examine the different perspectives on why transparency reports are important.

a) Legal perspective

Transparency reports are useful and often necessary to comply with EU regulations like the TCO Regulation, the Digital Services Act and country-specific laws where applicable. If HSPs have taken measures against the dissemination of terrorist content within a calendar year, whether proactively or in compliance with a removal order, a transparency report **must be published by 1 March of the following year** at latest (TCO Regulation, Art. 7.2). This means that transparency reports are more likely to be mandatory than not. Competent authorities are also required to publish annual transparency reports (TCO Regulation, 31).

b) User and stakeholder perspective

Transparency reports aid users and other stakeholders to assess the extent to which HSPs fulfil their **responsibility to society**. At the same time, regular publication of transparency reports helps the HSP to **build trust and a good public reputation** by demonstrating compliance, commitment, and reliability.

c) Company perspective

Similar to complaint mechanisms, transparency reports can also be a **self-monitoring tool** that identify areas where business processes can be optimised. Given that smaller HSPs are particularly popular with terrorist actors (for details, see this Tech Against Terrorism report), it is incumbent upon them to take action against the dissemination of terrorist content. Transparency reports are one way for smaller HSPs to demonstrate their engagement with this effort.

[9] Woolery, L., Budish, R., & Bankston, K. (2016). The transparency reporting toolkit. *New America and The Berkman Center for Internet & Society at Harvard University*.

3. A process for the preparation of transparency reports

The initial setup and delivery of a transparency report can be a daunting task. However, once a **process and routine for preparing annual transparency reports** have been established, there is typically no need to completely rework the structure of the report, and updates are typically sufficient. This guide will outline how to set up a transparency report from the HSP perspective, as well as what to consider in advance, during and after the creation process.

a) **Before** the creation of the transparency report

Scan the legal landscape

Get an overview of the legal regulations that apply to your HSP. For example, the TCO Regulation and the Digital Services Act are relevant throughout the EU. There may be country-specific regulations as well as requirements for other topics relevant to your HSP outside of terrorist content. If you have employed contacts or lawyers, it is often advisable to exchange with them in order to determine which regulations (apart from the TCO Regulation you are dealing with here) apply to your HSP.

Determine the objectives of the transparency report

Think about what the objectives of your transparency report are. Key questions that can help you with this are: Do you simply 'only' want to fulfil your legal obligations, or do you also want to address other topics and your commitment to them? Who do you want to reach, i.e. which target group makes sense for your HSP (e.g. political actors, users, financiers)? How often do you want to publish transparency reports and what is the best time for your individual financial year?

Determine which data can and should be included

Determine which data you are going to include in the transparency report. On the one hand, *ability* is relevant for this, i.e. what data is available or for which is it feasible that you can collect it in the future? On the other hand, the *must* is decisive, i.e.: what legal requirements do you have to meet and what data is necessary for this? You can find out what information and data you need to include in accordance with the TCO Regulation in ▶ the next subchapter.

b) **During** the creation of the transparency report

Use clear and concise language

While creating the transparency report, make sure that clear and concise language is used. This aids engagement with and comprehension of complex material. Language should be further adapted to the readership envisaged.

Provide contextual information and explanations

Provide your readers with contextual information and explanations. Such explanations enable users to better understand the way your HSP operates and the reasons as to why it (perhaps) provides a more detailed transparency report than 'just' a report meeting only the minimum regulatory requirements.

Incorporate internal feedback

When scheduling a transparency report, incorporate rounds of feedback into the production timeline. Regular feedback offering corrections to the report's content and language can be valuable for all actors involved, especially those tasked with preparing the report.

c) **After** the creation of the transparency report

Publish the transparency report

Think about the languages in which you want to publish the transparency report and create appropriate translations. Also determine where it should be accessible, i.e. where on your website. In addition, you can also consider including the transparency report in various communication materials to help it gain more attention. This could include embedding it on your website, sending it out in email newsletters, or sharing it on social media.

Regularly update the data and, ultimately, the transparency report

When the first transparency report has been completed, you have already built the foundation that will make the following one easier given that you have already created a template. The process for collecting relevant data throughout the year and organising it to make it readily retrievable will make the production of the next transparency report easier and less time- and resource-intensive.

Allow room for improvement

Be open to changes and adjustments. If you receive external feedback, consider incorporating it into the next transparency report if appropriate. However, relevant feedback may not only be external. After publication, HSP can also take a critical look at the previous transparency report, the communication around it and reactions to it, by considering press statements or other relevant sources. Both internal and external feedback can identify where there may be scope for improvement.

4. What information and metrics need to be included in the transparency report?

The TCO Regulation explains in Article 7.3 the **minimum requirements for transparency reports**, i.e. precisely what needs to be included to comply with this EU law. We present these requirements in a checklist below:

1) **Information** on what measures has the HSP taken:

- ☑ to identify terrorist content;
- ☑ to remove or disable terrorist content;
- ☑ to prevent the reappearance and reuploading of previously blocked online materials (this is particularly relevant when automated procedures are in use).

2) **Metrics** and, if applicable, additional information concerning the number of:

- ☑ removed items that include terrorist content (based on removal orders or other measures);
- ☑ removal orders that were not actioned and additional information on why this was not the case;
- ☑ complaints handled by the HSP through the complaint mechanism, as well as additional information on the outcome of the complaints;
- ☑ cases in which the HSP restored the content following the complaint of the content provider;
- ☑ legal proceedings initiated by the HSP and additional information on the outcome of these;
- ☑ cases in which the HSP has had to restore content following legal proceedings.

C. Thank You for Your Help to Counter the Terrorist Threat!

Congratulations!

You've made it this far, and that means you have gained essential knowledge on the requirements of the TCO Regulation and further measures to counter terrorist and other harmful content online. In this way, you will be contributing to a safer internet.

We are aware that the implementation of these measures requires considerable attention and resources. The fact that you have engaged with this guide is a great move and if this guide helps you to think about a strategy for implementing the measures against terrorist content that is suitable for your HSP, then we have already achieved a lot! We are confident that this guide , alongside our other educational resources ,will be an invaluable tool to enable your HSP to fully meet its obligations in terms of countering an online terrorist threat.

Thank you very much for your commitment to countering the online terrorist threat, and for standing up for your HSP and your users.

PS: The Tech Against Terrorism Europe project also includes a **free, awarded online course** to further explore compliance with the EU's TCO Regulation. In this course, you can delve deeper into the subject matter and find details on the regulation, examples of how other platforms implement individual measures, and more general information on terrorist behaviour online. **After successfully completing the course, you will receive an official certificate signed by renowned universities** (LMU Munich, University of Ghent). Additionally, TATE offers a Capacity Building Programme in which HSPs can receive hands-on-support for the requirements related to the TCO Regulation.

D. Glossary

Term	Explanation
Competent authorities	The authorities of an EU Member State that are responsible for implementing the TCO Regulation. An overview of the respective competent authorities of the Member States can be found here.
Content provider	The person who provides content on the respective platform, for example publishes a post.
HSP	Hosting service providers; the TCO Regulation applies to HSPs. More details about which HSPs are in the scope of the TCO Regulation can be found ▶here.
PERCI	PERCI is a tool coordinated by Europol that is intended to enhance and facilitate communication between HSPs and competent authorities.
Removal order	A request that a HSP receives from a competent authority. It informs the HSP that terrorist content has been circulated on the platform and obligates the HSP to swiftly remove it within one hour upon receipt.
TCO Regulation (also: LEX 2021/784 & Regulation on Addressing the Dissemination of Terrorist Content Online)	The EU regulation against the dissemination of terrorist content online that came into effect in 2022. It applies to hosting service providers (HSPs) that offer their services in the EU.
Terrorist Content Online (TCO)	Content that includes terrorist elements or is intended to promote terrorist purposes. A detailed definition of this can be found in the ▶introduction. Terrorist content online is closely related to ▶terrorist offences, which can be very diverse in nature.
ToS	Terms of Service; Binding rules established by and for the individual platform that (1) define the scope and responsibility of HSPs towards users and (2) appropriate and permitted, but also prohibited usage practices. Users must comply with them if they wish to continue using the service offered by HSPs. Numerous synonyms for ToS exist, e.g. Terms of Use, Terms and Conditions or Community Standards