tech against terrorism europe

Optimiser votre réponse antiterroriste

Un guide en six étapes pour appliquer la réglementation de l'UE relative à la lutte contre la diffusion des contenus à caractère terroriste en ligne

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

Finance par l'Union européenne

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

GHENT UNIVERSITY

Tech Against Terrorism Europe (TATE)

Ce guide fait partie du projet Tech Against Terrorism Europe (TATE). Le projet TATE est financé par l'Union européenne (ISF-2021-AG-CTEL-101080101). Ce projet soutient les petits fournisseurs de services d'hébergement (FSH) dans leur lutte contre le terrorisme et dans l'établissement de rapports de transparence, comme l'exigent le règlement de l'UE 2021/784 relatif à la lutte contre la diffusion des contenus à caractère terroriste en ligne et la directive de l'UE 2017/541.

tate.techagainstterrorism.org

Financé par l'Union européenne

Autrices

Sophia Rothut est doctorante dans le laboratoire de la professeure Diana Rieger (IfKW, Ludwig Maximilian University, Munich). Dans le cadre du projet européen TATE, elle travaille sur les exigences relatives à la lutte contre le contenu terroriste en ligne. Ses recherches portent sur la radicalisation en ligne, le mainstreaming et les influenceurs et influenceuses politiques d'extrême droite.

Heidi Schulze est doctorante dans le laboratoire de la professeure Diana Rieger au Département des médias et de la communication de la Ludwig Maximilian University de Munich. Dans le cadre du groupe de recherche interdisciplinaire MOTRA - Monitoring System and Transfer Platform Radicalization (BMBF & BMI), elle mène des recherches sur la radicalisation au sein des environnements en ligne, en se concentrant notamment sur les médias sociaux (alternatifs) et les sites web d'information à caractère idéologique.

Diana Rieger est professeure au Département des médias et de la communication de la Ludwig Maximilian University de Munich. Elle mène des recherches sur la radicalisation en ligne, les discours haineux et l'impact des contenus de divertissement sur la radicalisation. Elle se concentre également sur le développement et l'évaluation de contre-mesures.

Catherine Bouko est professeure de communication et de français à l'université de Gand (Belgique). Ses recherches portent principalement sur la communication politique, l'extrémisme et la citoyenneté sur les médias sociaux, avec un focus particulier sur la communication par l'image. Ses méthodes de recherche comprennent l'analyse (multimodale) du discours, l'analyse quantitative du contenu, la sémiotique et l'ethnographie.

Brigitte Naderer est chercheuse postdoctorale à l'Université médicale de Vienne au Centre de santé publique, Département de médecine sociale et préventive, Unité de recherche sur le suicide et la promotion de la santé mentale, et travaillait au Département des médias et de la communication (LMU Munich) jusqu'en mars 2023. Ses recherches portent sur l'éducation aux médias, la radicalisation en ligne et les effets des médias sur les enfants et les adolescents.

Remerciements : Nous remercions Tech Against Terrorism pour leur aide dans le cadre de l'élaboration de ce guide. Nous tenons également à remercier le professeur Maura Conway pour ses commentaires utiles.

tech against terrorism europe

Table des matières

A. Introduction 4

B. Principales obligations et recommandations relatives au règlement CTEL 8

Chapitre 1 Rédiger et appliquer des conditions d'utilisation interdisant les contenus terroristes 10

1. En quoi consistent les conditions d'utilisation ? 10

2. Pourquoi est-il nécessaire d'avoir des conditions d'utilisation claires et solides ? 11

3. Conseils pratiques : Quels sont les éléments importants pour des conditions d'utilisation solides ? 12

Chapitre 2 Etablir des mesures spécifiques d'identification et de suppression des contenus (terroristes) 15

1. Pourquoi est-il nécessaire de mettre en place des processus d'identification des contenus illicites et préjudiciables? 16

2. Un processus d'identification des contenus terroristes 16

3. Conseils pratiques : Qu'est-ce qui permet d'évaluer le caractère illicite d'un contenu ? 18

4. Que dois-je faire si je vois les choses différemment ? Comment contester une injonction de retrait reçue? 20

Chapitre 3 Mettre en place de mécanismes de modération efficaces pour les contenus terroristes en ligne 22

1. Qu'est-ce que la modération de contenu et pourquoi est-elle nécessaire dans certains cas ? 22

2. Conseils pratiques : Comment mettre en œuvre la modération de contenu ? 25

3. Autres approches de modération 27

Chapitre 4 Établir des points de contact et des représentants légaux 30

1. Que sont les points de contact et les représentants légaux ? 30

2. Pourquoi est-il nécessaire d'avoir un point de contact ou un représentant légal ? 31

3. Qu'est-ce que l'autorité compétente d'un État membre de l'UE et comment la contacter ? 32

Chapitre 5 Mettre en place d'un système de notification et de réclamation concernant les contenus supprimés 33

1. Pourquoi est-il nécessaire de mettre en place un mécanisme de réclamation transparent ? 33

2. Quelles sont les exigences d'un système de réclamation ? 34

3. Comment les plaintes sont-elles être traitées et quels sont les résultats possibles ? 34

4. Conseils pratiques : Quels sont les éléments utiles pour mettre en place un système de réclamation ? 35

Chapitre 6 Aide et conseils pratiques concernant les rapports de transparence 37

1. Qu'est-ce qu'un rapport de transparence ? 37

2. Pourquoi les rapports de transparence sont-ils nécessaires ? 38

3. Un processus de préparation des rapports de transparence 39

4. Quelles sont les informations et les mesures à inclure dans le rapport de transparence concernant le contenu terroriste en ligne (CTEL) ? 41

C. Merci de votre aide pour lutter contre la menace terroriste ! 42

D. Glossaire 43

A. Introduction

En quoi consiste ce guide et pourquoi vous devriez le lire ...

Ce guide traite des **obligations imposées aux fournisseurs de services d'hébergement (FSH) par le règlement européen sur les contenus à caractère terroriste en ligne (CTEL)**. Note : le sigle « CTEL » est utilisé dans ce guide (pour « contenu terroriste en ligne »). Ce guide aborde les éléments que les fournisseurs de services d'hébergement doivent prendre en compte pour lutter contre la diffusion de contenus terroristes en ligne et donne des conseils pratiques sur la manière de mettre en œuvre diverses mesures.

Le guide se concentre tout d'abord sur les **exigences minimales que les fournisseurs de services d'hébergement doivent respecter pour se conformer au règlement CTEL**. Ensuite, il fournit des **conseils pratiques sur les mesures (proactives) pertinentes que les FSH sont encouragés à adopter** pour appliquer la réglementation (par exemple, que faire lorsque vous recevez une injonction de retrait) et pour préparer les FSH à contrer l'exploitation à caractère terroriste de leurs plates-formes.

Ce guide a été élaboré dans le cadre du projet Tech Against Terrorism Europe (TATE) financé par la Commission européenne. Le consortium TATE est composé de chercheurs et chercheuses des institutions suivantes : Dublin City University, Ghent University, JOS Project, LMU Munich, Saher Europe, Swansea University et Tech Against Terrorism. TATE vise à faire connaître le règlement CTEL de l'UE et à aider les petites et microentreprises technologiques à prendre des mesures visant à appliquer le règlement CTEL. TATE a créé ce guide et d'autres ressources pour rendre les attentes juridiques et pratiques aussi concrètes que possibles pour les fournisseurs de services d'hébergement.

À qui s'adresse ce guide ?

Ce guide s'adresse aux fournisseurs de services d'hébergement (FSH) et à leurs employés et employées, ainsi qu'aux professionnels et professionnelles de l'informatique qui souhaitent mettre en œuvre des dispositifs techniques de lutte contre les contenus terroristes sur leurs plateformes. Il vise à fournir des informations sur les exigences minimales à respecter pour se conformer au règlement CTEL de l'UE de 2021 (29 avril).

Nous avons créé ce guide et d'autres ressources TATE pour soutenir les petites et micro-entreprises technologiques. Nous sommes conscientes que celles-ci disposent souvent de ressources limitées pour contrer les utilisations à caractère terroriste de leurs plateformes. **Toutefois, il est essentiel que les petits et micro FSH ne négligent pas la menace, car les petites plateformes sont particulièrement susceptibles d'être exploitées par des acteurs terroristes** (pour plus de détails, voir le rapport de Tech Against Terrorism).

Quels sont les principaux éléments du règlement de l'UE sur le contenu terroriste en ligne (CTEL) ?

Le règlement européen sur les contenus terroristes en ligne (CTEL) est entré en vigueur en juin 2022 et **oblige les fournisseurs de services d'hébergement à supprimer les contenus ou à en désactiver l'accès dans l'heure qui suit la réception d'une injonction de retrait émanant d'une autorité compétente**.

Les fournisseurs de services d'hébergement doivent coopérer avec les autorités chargées de l'application de la loi, telles qu'Europol, ainsi qu'avec les autres autorités compétentes, pour détecter et supprimer les contenus terroristes susceptibles d'être présents sur leurs plates-formes. Pour se conformer au règlement CTEL, les **fournisseurs de services d'hébergement doivent également mettre en œuvre des mesures efficaces et proportionnées pour empêcher que les contenus terroristes soient réimportés** et sont soumis à d'autres obligations que vous découvrirez dans ce guide.

Comment ce guide est-il structuré ?

Le règlement CTEL exige actuellement que les fournisseurs de services d'hébergement :

1) Rédigent des conditions d'utilisation appropriées (▶chapitre 1),
2) Prennent des mesures spécifiques pour identifier et supprimer le contenu terroriste (▶chapitre 2),
3) Mettent en place des mécanismes de modération efficaces (▶chapitre 3),
4) Établissent des points de contact et des représentants légaux (▶chapitre 4),
5) Mettent en place des mécanismes de notification et de réclamation pour les utilisateurs et utilisatrices (▶chapitre 5),
6) Rédigent des rapports de transparence (▶chapitre 6).

Nous avons structuré les explications et les recommandations fournies dans ce guide en fonction de ces six exigences principales. Au début de chaque chapitre, vous trouverez un bref résumé du contenu du chapitre et de ses principaux points. Ensuite, vous trouverez des informations détaillées sur ce que le règlement CTEL exige ou recommande, ainsi que des conseils pratiques pour défendre votre plateforme contre les contenus terroristes (et autres contenus préjudiciables).

Quelles sont les plateformes concernées par le règlement CTEL ?

Le règlement CTEL concerne les fournisseurs de services d'hébergement, c'est-à-dire toute plate-forme permettant aux utilisateurs de diffuser des informations au public par l'intermédiaire de ses services.

Le règlement CTEL s'applique aux **fournisseurs de services d'hébergement de toutes tailles offrant leurs services dans l'UE**. Il s'applique également aux FSH situés en dehors de l'UE lorsqu'un FSH 1) a un nombre significatif d'utilisateurs et utilisatrices dans un ou plusieurs États membres de l'UE ou 2) cible ses activités dans un ou plusieurs États membres de l'UE.

Quand un FSH est-il considéré comme « exposé à des contenus à caractère terroriste » ?

Le règlement CTEL impose des obligations spécifiques aux fournisseurs de services d'hébergement qui sont « exposés à des contenus à caractère terroriste ». Selon l'art. 5.4, une telle exposition se produit lorsqu'un FSH a été notifié et a reçu au moins deux injonctions de retraits définitives au cours des 12 derniers mois de la part de l'autorité compétente de l'État membre dans lequel le FSH a son établissement principal ou son représentant légal dans l'UE.

Qu'est-ce que du contenu à caractère terroriste ?

Le règlement CTEL portant sur les contenus à caractère terroriste diffusés en ligne, il est essentiel d'en donner une définition. La directive 2017/541 de l'UE jette les bases du règlement CTEL en définissant ce que l'on entend par « contenu terroriste ».

Un contenu est considéré comme terroriste lorsqu'il incite à commettre des actions ou nourrit des intentions en faveur d'une cause terroriste, contribuant ainsi directement ou indirectement à la menace d'infractions terroristes.

Menacer de commettre une infraction terroriste est également considéré comme un contenu terroriste, de même que fournir des informations, soutenir ou financer de tels actes.

Types d'infractions terroristes

Les infractions terroristes peuvent inclure les éléments suivants :

(directive 2017/541 de l'UE, article 3.1)

- Les atteintes à la vie ou à l'intégrité physique d'une personne
- L'enlèvement ou la prise d'otage
- La destruction massive d'installations et d'infrastructures spécifiques (par exemple, installations gouvernementales/publiques, systèmes de transport et d'information) susceptible de mettre en danger des vies humaines ou d'entraîner des pertes économiques majeures.
- La saisie d'aéronefs, de navires ou d'autres moyens de transport public ou de marchandises.
- La fabrication, la détention, l'acquisition, le transport, la fourniture ou l'utilisation d'explosifs ou d'armes et recherche et le développement d'armes chimiques, biologiques, radiologiques ou nucléaires.
- Le rejet de substances dangereuses, les incendies, les inondations ou les explosions, ainsi que la perturbation des ressources fondamentales (par exemple, l'eau, l'électricité) mettant en danger la vie humaine.
- La perturbation ou l'interruption de l'approvisionnement en eau, en électricité ou en toute autre ressource naturelle fondamentale, par laquelle la vie humaine est mise en danger.

Le fait de diriger un groupe terroriste ou de participer intentionnellement à ses activités est également considéré comme une infraction pénale (directive 2017/541 de l'UE, article 4). Cela inclut la fourniture de ressources (informationnelles) (par exemple, des instructions ou des matériaux pour fabriquer des armes) pour des causes terroristes ou le financement de telles activités.

Au moyen de ces infractions, les acteurs terroristes visent à a) intimider gravement le public, b) contraindre indûment un gouvernement ou une organisation internationale à (s'abstenir de) réaliser un acte spécifique, ou c) déstabiliser gravement ou détruire les structures fondamentales politiques, constitutionnelles, économiques ou sociales d'un pays ou d'une organisation internationale (directive 2017/541 de l'UE, art. 3.2).

En résumé, un contenu est considéré comme terroriste lorsqu'il permet, soutient ou facilite une infraction terroriste ou lorsqu'il contient une menace de commettre une infraction terroriste.

B. Principales obligations et recommandations relatives au règlement CTEL

Dans les chapitres suivants, vous trouverez des informations sur six éléments et étapes-clés pour adapter votre plate-forme à la menace terroriste. Il s'agit des domaines et questions-clés suivants.

Chapitre 1 : Rédiger et appliquer des conditions d'utilisation interdisant les contenus terroristes

1. En quoi consistent les conditions d'utilisation ?
2. Pourquoi est-il nécessaire d'avoir des conditions d'utilisation claires et solides ?
3. Conseils pratiques : Quels sont les éléments importants pour des conditions d'utilisation solides ?

Chapitre 2 : Etablir des mesures spécifiques d'identification et de suppression des contenus (terroristes)

1. Pourquoi est-il nécessaire de mettre en place des processus d'identification des contenus illicites et préjudiciables ?
2. Un processus d'identification du contenu terroriste
3. Conseils pratiques : Qu'est-ce qui permet d'évaluer le caractère illicite d'un contenu ?
4. Que dois-je faire si je vois les choses différemment ? Comment contester une injonction de retrait reçue ?

Chapitre 3 : Mettre en place des mécanismes de modération efficaces pour les contenus terroristes en ligne

1. Qu'est-ce que la modération de contenu et pourquoi est-elle nécessaire dans certains cas ?
2. Conseils pratiques : Comment mettre en œuvre la modération de contenu ?
3. Autres approches de modération

Chapitre 4 : Établir des points de contact et des représentants légaux

1. Que sont les points de contact et les représentants légaux ?
2. Pourquoi est-il nécessaire d'avoir un point de contact ou un représentant légal ?
3. Qu'est-ce que l'autorité compétente d'un État membre de l'UE et comment la contacter ?

Chapitre 5 : Mettre en place un système de notification et de réclamation concernant les contenus supprimés

1. Pourquoi est-il nécessaire de mettre en place un mécanisme de réclamation transparent ?
2. Quelles sont les exigences d'un système de réclamation ?
3. Comment les plaintes sont-elles traitées et quels sont les résultats possibles ?
4. Conseils pratiques : Quels sont les éléments utiles pour mettre en place un système de réclamation ?

Chapitre 6 : Aide et conseils pratiques concernant les rapports de transparence

1. Qu'est-ce qu'un rapport de transparence ?
2. Pourquoi les rapports de transparence sont-ils nécessaires ?
3. Un processus de préparation des rapports de transparence
4. Quelles sont les informations et les mesures à inclure dans le rapport de transparence concernant le contenu terroriste en ligne (CTEL) ?

Le guide explique les **aspects du règlement CTEL les plus pertinents pour les plateformes**, mais **il fournit également des conseils pratiques sur les mesures proactives qui peuvent être prises pour lutter contre la diffusion de contenus préjudiciables en ligne**. Ces mesures sont essentielles pour aider les plateformes à lutter contre l'exploitation terroriste de leurs services ou, dans le cadre du règlement CTEL, pour être en mesure de faire face à la situation lorsque la première injonction de retrait et toute injonction ultérieure atterrissent dans votre boîte mails.

Chapitre 1

Rédiger et appliquer des conditions d'utilisation interdisant les contenus terroristes

Résumé : Contenu et points principaux de ce chapitre

- Les conditions d'utilisation sont un **accord contraignant** entre l'utilisateur/l'utilisatrice et le FSH, qui définit l'utilisation appropriée et autorisée de sa plate-forme.
- Les fournisseurs de services d'hébergement doivent (1) définir leur **stratégie de lutte contre la diffusion de contenus terroristes** dans leurs conditions d'utilisation et (2) **interdire la** diffusion de **contenus terroristes.**
- En outre, au-delà du champ d'application du règlement CTEL, les fournisseurs de services d'hébergement peuvent et sont encouragés à envisager d'interdire d'autres formes de contenus préjudiciables (par exemple, les contenus extrémistes, les discours de haine).
- Les conditions générales d'utilisation sont une **nécessité juridique** dans le cadre du règlement CTEL et peuvent également constituer des **protections utiles pour la plateforme.**
- Plusieurs éléments sont nécessaires pour rendre vos conditions d'utilisation aussi solides que possible.

Les conditions d'utilisation vous aident à **définir les bases de l'interdiction et, par conséquent, du traitement des contenus terroristes.** Le règlement CTEL invite explicitement les fournisseurs de services d'hébergement à exposer « leur politique de lutte contre la diffusion des contenus à caractère terroriste, accompagnée, le cas échéant, d'une explication pertinente du fonctionnement des mesures spécifiques, y compris, s'il y a lieu, du recours à des outils automatisés. » (règlement CTEL, article 7.1).

En outre, et au-delà du champ d'application du règlement CTEL, les fournisseurs de services d'hébergement peuvent et sont encouragés à **envisager d'interdire d'autres formes de contenus préjudiciables** dans leurs conditions d'utilisation (par exemple, les contenus extrémistes, les discours de haine, l'apologie de la violence.

1. En quoi consistent les conditions d'utilisation ?

Les conditions d'utilisation sont des règles établies par et pour la plateforme spécifique qui définissent (1) les responsabilités du FSH envers ses utilisateurs et utilisatrices, et (2) le comportement et le contenu appropriés et autorisés, mais aussi interdits, sur cette plateforme. Les utilisateurs et utilisatrices doivent

accepter ces conditions s'ils et elles veulent utiliser le service fourni par le fournisseur de services d'hébergement.

Les synonymes utilisés pour les conditions d'utilisations sont, par exemple, les conditions générales ou les standards de la communauté. Le règlement CTEL comprend l'expression « conditions générales » et la définit comme « l'ensemble des modalités, conditions et clauses, indépendamment de leur dénomination ou de leur forme, qui régissent la relation contractuelle entre un fournisseur de services d'hébergement et ses utilisateurs » (règlement CTEL, article 2.8).

2. Pourquoi est-il nécessaire d'avoir des conditions d'utilisation claires et solides ?

Pour mieux comprendre pourquoi il est nécessaire d'avoir des conditions d'utilisation claires et solides, nous adoptons deux points de vue différents : le point de vue juridique et le point de vue de l'entreprise/opérationnel.

a) Point de vue juridique

Conformité avec la législation (de l'UE)

Les conditions générales doivent être conformes au droit (de l'UE) en vigueur, notamment le règlement CTEL mais aussi la directive 2017/541 de l'UE, qui renforce la lutte contre les contenus terroristes à l'échelle de l'UE (1) en fournissant une définition des contenus terroristes, (2) en établissant des sanctions à leur encontre (article 15), (3) en renforçant les droits et le soutien des victimes et (4) en reconnaissant le terrorisme comme une menace transnationale et transfrontalière. La directive et le règlement CTEL encouragent tous deux la coopération européenne et internationale. Les conditions d'utilisation sont un moyen pour les fournisseurs de services d'hébergement de souligner leur engagement dans la lutte contre les activités terroristes.

Protection du FSH contre la responsabilité juridique

Les conditions générales ont pour but de protéger le FSH de toute responsabilité (pour autant qu'elle soit conforme à la législation applicable). Étant donné qu'il s'agit d'un contrat entre le FSH et ses utilisateurs et utilisatrices, les conditions générales constituent un accord juridiquement contraignant entre ces deux parties. Elles permettent au FSH de définir des règles d'utilisation (non) acceptables en accord avec les valeurs du FSH.

Justification du retrait de contenu (proactif)

En outre, des conditions d'utilisation claires sont importantes lorsqu'il s'agit de faire face à des contenus problématiques et de les modérer. Les fournisseurs de services d'hébergement peuvent se référer à leurs conditions générales pour justifier la modération de contenu si elles contiennent des informations détaillées sur les pratiques interdites en matière de contenu, y compris l'interdiction du terrorisme.

b) Point de vue de l'entreprise

Maintien de l'activité opérationnelle

Tout d'abord, le respect de la législation est essentiel à la poursuite de l'activité opérationnelle et, par conséquent, au succès de l'entreprise. Les fournisseurs de services d'hébergement ont une certaine responsabilité vis-à-vis de leurs actionnaires (le cas échéant), qui s'attendent à ce qu'ils respectent la législation. En outre, la responsabilité financière peut être engagée, et les sanctions réglementaires ou l'atteinte à la réputation de l'entreprise sont d'autres conséquences possibles en cas de non-conformité.

Mise en valeur de la responsabilité publique du FSH

Deuxièmement, en tant qu'acteurs clés du monde numérique, les FSH ont une responsabilité à l'égard du public, tant d'un point de vue collectif qu'individuel. La responsabilité envers le public en général implique de contrer les activités illégales telles que le terrorisme et de détecter et prévenir les contenus terroristes avant qu'ils ne soient largement diffusés. En indiquant dans les conditions générales que les activités illégales sont interdites, les fournisseurs de services d'hébergement posent les bases d'un environnement en ligne plus sûr et peuvent s'appuyer sur ces conditions d'utilisation lorsqu'ils prennent des mesures à l'encontre de contenus illégaux. Des conditions générales claires aident également les entreprises à démontrer leur responsabilité envers les membres du public en instaurant la confiance grâce à leur transparence. À cet égard, les fournisseurs de services d'hébergement montrent aux utilisateurs et utilisatrices qu'ils et elles seront protégés contre les contenus préjudiciables sur la plateforme.

3. Conseils pratiques : Quels sont les éléments importants pour des conditions d'utilisation solides ?

Les sections suivantes fournissent des conseils sur les éléments que vous, en tant que fournisseur de services d'hébergement, devez prendre en compte lors de la rédaction des conditions d'utilisation dans le contexte du règlement CTEL.

Il est important de souligner que l'élaboration des conditions générales est un processus itératif : elles peuvent et doivent être adaptées si nécessaire, par exemple en réponse aux suggestions des utilisateurs et utilisatrices et en suivant l'évolution de la législation nationale ou internationale. Le règlement CTEL préconise une telle adaptabilité : les fournisseurs de services d'hébergement, s'ils sont exposés à des contenus terroristes, sont tenus de modifier leurs conditions générales pour indiquer que des **mesures seront prises pour lutter contre l'utilisation abusive de la plateforme à des fins terroristes** (règlement CTEL, article 5.1 1).

En particulier, nous recommandons de prêter attention à la **clarté et à la structure** des conditions générales. Des études montrent que les utilisateurs et utilisatrices consacrent peu de temps à la lecture des conditions générales et traitent les informations présentées avec peu de concentration ; les conditions générales

permettent de sensibiliser les utilisateurs et utilisatrices à ces aspects juridiques[1,2]. Pour faciliter la navigation et la compréhension des conditions générales, il est conseillé d'investir du temps dans leur conception visuelle et de travailler avec des titres clairs ou des listes à puces. Il peut également être utile de rappeler aux utilisateurs et utilisatrices certains éléments des conditions générales dans le cadre d'une utilisation quotidienne ou, si un comportement suspect a été détecté, de les avertir des conséquences possibles au moyen d'une fenêtre pop-up[2].

Lors de l'élaboration et de la révision des conditions générales, la check-list suivante **peut servir de guide pour optimiser leur conformité avec le règlement CTEL**.

Définir le contenu terroriste

Il est important de disposer d'une définition pratique du terrorisme et du contenu terroriste afin de pouvoir évaluer les contenus et les comportements en fonction de cette définition. Cette définition doit être mentionnée dans les conditions générales et l'évaluation du contenu doit y faire référence. Les définitions existantes déjà présentées et discutées dans ce guide (►voir la définition ici), en particulier celles de l'UE basées sur la directive 2017/541 de l'UE, peuvent fournir des orientations utiles à cet égard.

Expliciter la stratégie de lutte contre les contenus terroristes

Le règlement CTEL oblige les FHS exposés à des contenus terroristes à expliquer dans leurs conditions générales à la fois la stratégie avec laquelle ils entendent lutter contre la diffusion de contenus terroristes et tout moyen automatisé utilisé (par exemple, pour identifier les contenus interdits ; ►chapitre 2). Le règlement CTEL exige en outre des fournisseurs de services d'hébergement qu'ils mettent en œuvre des mesures spécifiques lorsqu'ils sont exposés à des contenus terroristes (par exemple, des systèmes de signalement de contenus préjudiciables par les utilisateurs et utilisatrices, des systèmes de réclamation dans le cas de la suppression de contenus ; ►chapitre 4) (règlement CTEL, art. 7.1).

Utiliser des répertoires et envisager d'interdire clairement d'autres catégories de contenus préjudiciables

De multiples formes de communication peuvent être préjudiciables aux utilisateurs et utilisatrices, à la société et à la plateforme elle-même. Les fournisseurs de services d'hébergement devraient donc prendre en considération les interdictions supplémentaires suivantes, également prises en compte par d'autres cadres juridiques, et donc interdire les contenus contenant, par exemple, des discours de haine ou des incitations à la violence. Au niveau des acteurs, les groupes terroristes peuvent être exclus de l'utilisation des plateformes. Les répertoires (c'est-à-dire les listes officielles de groupes terroristes publiées par les États démocratiques) constituent une bonne ressource pour

[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. et 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

de telles interdictions. En outre, les autorités compétentes peuvent alerter les fournisseurs de services d'hébergement sur des contenus qui ne sont pas classés comme terroristes mais qui sont néanmoins préjudiciables ; l'évaluation du contenu par rapport aux conditions générales et le traitement éventuel du contenu non terroriste mais préjudiciable relèvent en dernier ressort de la responsabilité des fournisseurs de services d'hébergement (règlement CTEL, 40).

Expliciter les usages acceptés

Les conditions générales doivent également décrire, en lien avec les valeurs de la plateforme, les comportements d'utilisation qui sont non seulement acceptés, mais aussi accueillis et encouragés par le FSH. Il n'existe pas de norme universelle en la matière : cela varie considérablement en fonction des fonctionnalités et des valeurs de la plateforme. Néanmoins, la description des usages acceptés n'est pas explicitement requise pour satisfaire aux obligations du règlement CTEL.

Fournir des informations sur la manière de signaler les contenus interdits

Les systèmes de signalement par les utilisateurs et utilisatrices constituent un moyen d'identifier les contenus illégaux et les contenus enfreignant les conditions générales. Les fournisseurs de services d'hébergement doivent donc inclure dans leurs conditions générales une section expliquant les procédures de signalement des contenus suspectés de les enfreindre.

Déterminer les conséquences d'une violation des conditions générales

Pour s'assurer que les violations des conditions générales sont traitées de manière transparente et pour prendre toutes leurs responsabilités envers le public, les fournisseurs de services d'hébergement doivent clairement définir dans les conditions générales les mesures à prendre à l'encontre d'un utilisateur/d'une utilisatrice ou d'un contenu en cas de violation des conditions générales. S'il s'agit d'une injonction de retrait émanant de l'autorité compétente, son traitement est incontestable : le contenu doit être retiré dans un délai d'une heure. Les injonctions de retrait peuvent également être suivis d'effets en bloquant le contenu ou en le géobloquant dans l'UE.

Publier des rapports annuels de transparence

Lors de la préparation et de la révision des conditions générales, les FSH doivent s'engager publiquement à publier régulièrement des rapports de transparence. Des conseils pratiques et les exigences réglementaires du CTEL pour la création de rapports de transparence figurent au ►chapitre 6.

Chapitre 2

Etablir des mesures spécifiques d'identification et de suppression des contenus (terroristes)

Résumé : Contenu et points principaux de ce chapitre

- La capacité des fournisseurs de services d'hébergement à identifier le contenu terroriste est fondamentale pour la mise en œuvre du règlement CTEL et au-delà.
- Dans de nombreux cas, les décisions à prendre ne sont pas faciles, peuvent être **très contextuelles et sensibles**, et les **droits fondamentaux tels que la liberté d'expression doivent être soigneusement pris en considération.**
- **En ce qui concerne le règlement CTEL, les fournisseurs de services d'hébergement n'ont pas à évaluer la légalité des contenus signalés par une autorité compétente**. Cependant, une fois qu'une injonction de retrait a été émise, les **fournisseurs de services d'hébergement et les fournisseurs de contenu (c'est-à-dire les utilisateurs et utilisatrices qui ont publié ce contenu) ont le droit de le contester**.
- Il est conseillé de mettre en place un processus d'identification des contenus terroristes. Ce processus comprendra plusieurs étapes.
- Diverses considérations pratiques, telles que l'utilisation de répertoires ou de bases de données reprenant les symboles ainsi qu'une réflexion sur les méthodes permettant de modérer les contenus, sont à prendre en compte.
- En outre, les procédures de réexamen jouent un rôle particulièrement important dans le cas particulier des injonctions de retrait **transfrontalières**, lorsqu'un FSH ne reçoit pas l'injonction de l'autorité compétente du pays dans lequel il a son principal établissement ou son représentant légal. **Des procédures spéciales s'appliquent** dans ce cas.

Il est fondamental que les fournisseurs de services d'hébergement soient en mesure d'évaluer si un contenu est illégal ou terroriste afin de le détecter de manière proactive et de contester les mesures de suppression s'ils estiment qu'elles ont été prises à tort. Il est important d'identifier correctement les contenus illégaux et terroristes afin de préserver les contenus légaux et non terroristes en ligne. Ces décisions ne sont souvent pas faciles à prendre et il est **toujours nécessaire d'évaluer le préjudice causé par le contenu et le préjudice causé par la violation du droit fondamental à la liberté d'expression**.

1. Pourquoi est-il nécessaire de mettre en place des processus d'identification des contenus illicites et préjudiciables?

Les processus d'identification comprennent une série définie d'étapes que l'entreprise ou la plateforme entreprend pour déterminer si un contenu enfreint le règlement CTEL ou des éléments des de ses conditions générales. Ces étapes permettent une évaluation méthodique et une conclusion motivée sur l'illégalité d'un contenu.

Ces décisions sont **indispensables pour** (a) **prendre des mesures proactives** contre les contenus terroristes et/ou autrement préjudiciables et (b) **contester les cas litigieux d'injonctions de retrait** émises par l'autorité compétente sur la base du règlement CTEL (de plus amples informations sur la procédure de contestation suivront dans le ►sous-chapitre 4).

Des mesures proactives contre les contenus terroristes et autres contenus préjudiciables sont recommandées pour diverses raisons.

1. En mettant en œuvre de telles mesures, les FSH font preuve de fiabilité et de bonne volonté à l'égard des décideurs et décideuses politiques, des autorités compétentes et des autres parties prenantes.
2. Les mesures proactives renforcent la confiance des parties prenantes car elles limitent positivement la responsabilité juridique ou l'atteinte à la réputation qui pourrait résulter de l'exploitation de la plateforme par des terroristes.
3. Une action insuffisante ou inadéquate contre les contenus terroristes, extrémistes ou autrement préjudiciables peut avoir pour effet d'attirer sur la plateforme des utilisateurs et utilisatrices qui ont l'intention de diffuser ce type de contenu. Il en résulterait une prévalence accrue des contenus interdits, ce qui, à son tour, entraînerait des investissements plus importants en temps, en efforts et en ressources financières de la part du fournisseur de services d'hébergement pour y faire face.

2. Un processus d'identification des contenus terroristes

1) Définition du contenu terroriste

Toute classification de contenu se base sur une définition. En ce qui concerne le règlement CTEL, la définition du contenu terroriste est celle contenue dans la directive 2017/541 de l'UE, comme mentionné dans l'►introduction.

Elle définit le contenu terroriste comme suit :

> « Le matériel qui incite ou invite quelqu'un à commettre des infractions terroristes ou à contribuer à la commission de telles infractions, invite quelqu'un à participer aux activités d'un groupe terroriste ou glorifie les activités terroristes y compris en diffusant du matériel représentant une attaque terroriste. La définition devrait également englober le matériel qui fournit des instructions concernant la fabrication ou l'utilisation d'explosifs, d'armes à feu ou d'autres armes, ou de substances nocives ou dangereuses, ainsi que de substances chimiques, biologiques, radiologiques et nucléaires (CBRN), ou concernant d'autres méthodes ou techniques spécifiques, y compris le choix de cibles, aux fins de la commission ou de la contribution à la commission d'infractions terroristes ». (règlement CTEL, 11)

La définition n'inclut pas « le matériel diffusé à des fins éducatives, journalistiques, artistiques ou de recherche ou à des fins de sensibilisation contre les activités terroristes » (règlement CTEL, (12)), et les droits fondamentaux tels que la liberté d'expression, d'information et de science doivent toujours être soigneusement pris en considération. « En outre, l'expression d'opinions radicales, polémiques ou controversées dans le cadre du débat public sur des questions politiques sensibles ne devrait pas être considérée comme étant un contenu à caractère terroriste ». (règlement CTEL, (12))

2) Optimisation grâce aux outils automatisés

Il peut souvent être très utile d'utiliser des outils automatisés pour l'identification préliminaire des contenus potentiellement problématiques. Par exemple, l'automatisation peut être utilisée pour identifier des mots-clés spécifiques associés à des visions extrémistes du monde ou à des éléments visuels tels que des logos ou des symboles d'organisations terroristes. Il existe également des initiatives visant à détecter automatiquement les contenus potentiellement terroristes sur les différentes plateformes en se référant à une base de données commune (par exemple, la plateforme d'analyse des contenus terroristes [TCAP] de Tech Against Terrorism ou la base de données de partage de hashs gérée par le Global Internet Forum to Counter Terrorism [GIFCT]).

3) Aide des systèmes de signalement

Si un FSH est exposé à un contenu terroriste - ce qui signifie formellement que l'autorité compétente du pays hébergeant l'activité principale du FSH ou son représentant légal a émis au moins deux injonctions de retrait juridiquement contraignantes et que le FSH en a été informé - le règlement CTEL exige explicitement que des mesures soient prises pour empêcher la diffusion de contenu terroriste (règlement CTEL, article 5). Toutefois, des mesures préventives sont également recommandées. L'une de ces mesures peut être un système de signalement qui permet aux utilisateurs de signaler au fournisseur de services d'hébergement les contenus suspects et interdits. Il peut être utile de demander aux utilisateurs et utilisatrices de spécifier la catégorie qu'ils attribueraient au contenu suspect. Le terrorisme devrait faire l'objet d'une catégorie distincte. Cela permet aux employés qui gèrent ces notifications de les hiérarchiser et de traiter plus rapidement les contenus soupçonnés de propager des idées terroristes. Cette pré-catégorisation peut également faciliter la préparation de rapports de transparence ultérieurs - plus d'informations à ce sujet au ▶ chapitre 6.

4) Examen par des modérateurs et modératrices humains

Le signalement de contenus par des outils automatisés ou par des utilisateurs et utilisatrices individuels est la première étape pour identifier les contenus suspects et potentiellement dangereux. Par la suite, cependant, un examen par des modérateurs humains est généralement nécessaire afin de prendre une décision sur les mesures à appliquer. Ces modérateurs et modératrices doivent (1) bien connaître la réglementation applicable à la plateforme, (2) être en mesure de différencier les contenus interdits des contenus qui mettent en danger le droit à la liberté d'expression, et (3) avoir une connaissance précise des différentes options de modération sur la plateforme. Il est important que les modérateurs humains soient constamment formés aux stratégies terroristes en ligne, mais il est tout aussi important, d'un point de vue éthique, qu'ils soient conseillés sur les effets psychologiques du traitement des contenus problématiques.

5) Décision sur la manière de traiter le contenu

Le processus décrit ci-dessus aboutit à une décision sur la manière de traiter le contenu. Il se peut que rien ne soit fait et que le contenu reste actif si, une fois tous les éléments pris en compte, il est jugé inoffensif. À l'autre extrême, dans le cas d'un contenu très problématique, dangereux ou interdit, tel que des appels à des attaques terroristes, le contenu et les utilisateurs/utilisatrices peuvent être bloqués. Entre les deux, il existe de nombreuses autres façons de traiter les contenus. Les contenus qui ne relèvent pas du règlement CTEL peuvent nécessiter un traitement plus nuancé. Vous trouverez des détails et des suggestions pratiques dans le ▶troisième chapitre sur les stratégies de modération efficaces.

6) Notification du fournisseur/de la fournisseuse de contenu

Si un contenu est bloqué, la personne qui a fourni ce contenu doit en être informée. Pour ce faire, il est recommandé de mettre en place un système de notification et de réclamation automatique. Vous trouverez des détails et des suggestions sur la manière de procéder en conformité avec le règlement CTEL dans le ▶cinquième chapitre.

3. Conseils pratiques : Qu'est-ce qui permet d'évaluer le caractère illicite d'un contenu ?

En ce qui concerne le règlement CTEL, les fournisseurs de services d'hébergement ne sont pas tenus d'évaluer la légalité du contenu, car il incombe aux autorités compétentes de le faire avant d'émettre une injonction de retrait. Toutefois, les **fournisseurs de services d'hébergement sont tenus d'identifier de manière proactive les contenus terroristes sur leurs plateformes une fois qu'ils ont été exposés à des contenus terroristes**. Il est souvent difficile de déterminer si un contenu franchit réellement la ligne légale et doit donc être interdit. En outre, il ne faut pas négliger le stress psychologique causé par le visionnage de contenus problématiques. Les employés et employées qui modèrent les contenus devraient avoir régulièrement l'occasion de prendre du recul par rapport à leur travail et, si nécessaire, de bénéficier d'un soutien psychologique.

Pour évaluer l'illégalité d'un contenu, vous trouverez ci-dessous quelques indicateurs et conseils pratiques. Avant cela, examinons les parties relatives à l'évaluation du contenu dans le règlement CTEL:

Le paragraphe 11 stipule, entre autres, que

> « Lorsqu'ils évaluent si un matériel constitue un contenu à caractère terroriste au sens du présent règlement, les autorités compétentes et les fournisseurs de services d'hébergement devraient tenir compte de facteurs tels que la nature et la formulation de déclarations, le contexte dans lequel les déclarations ont été faites et le fait qu'elles soient susceptibles d'engendrer des conséquences préjudiciables pour la sécurité et la sûreté des personnes. » (règlement CTEL, 11)

En outre, les droits fondamentaux doivent toujours être mis en balance :

> « Pour déterminer si le matériel fourni par un fournisseur de contenus constitue un 'contenu à caractère terroriste' au sens du présent règlement, il convient de tenir compte en particulier du droit à la liberté d'expression et d'information, y compris la liberté et le pluralisme des médias, et à la liberté des arts et des sciences. En particulier dans les cas où le fournisseur de contenus assume une

responsabilité éditoriale, toute décision relative au retrait du matériel diffusé devrait tenir compte des normes journalistiques établies par la réglementation de la presse ou des médias conformément au droit de l'Union, y compris la Charte. » (règlement CTEL, 12)

Si les fournisseurs de services d'hébergement reçoivent une injonction de retrait, le contenu a déjà été classé comme terroriste par l'autorité compétente. Dans **deux cas de figure**, il est utile que les fournisseurs de services d'hébergement soient compétents pour **pouvoir classer un contenu** comme terroriste ou non terroriste :

1. Si un fournisseur de services d'hébergement n'a pas encore reçu d'injonction de retrait officielle mais souhaite **agir de manière proactive** et/ou faire respecter ses propres conditions d'utilisation en ce qui concerne les contenus suspects.
2. Si un FSH reçoit une injonction de retrait officielle mais a **des doutes** sur l'évaluation du contenu en tant que terroriste par l'autorité compétente et souhaite faire réexaminer la mesure (en tant que première étape vers un recours juridique).

Vous trouverez ci-dessous diverses formes d'aide pratique pour classer un contenu comme terroriste ou non terroriste.

VOUS VOULEZ EN FAIRE PLUS POUR ASSURER LA SÉCURITÉ DE VOTRE PLATE-FORME ?

Utiliser des répertoires

Les répertoires nationaux et internationaux qui désignent les organisations terroristes constituent un bon cadre et un point de référence pour classer les contenus. Par exemple, l'adoption de répertoires comme base pour l'interdiction ou le blocage, qui devrait être référencée dans les conditions générales (▶chapitre 1), donne aux sanctions d'une plateforme le soutien de la loi. Le paragraphe (11) du règlement CTEL suggère explicitement que la liste de l'Union européenne peut être utilisée lors de l'évaluation du contenu.

Utiliser les bases de données de mots-clés, de symboles et de logos

Les personnes qui examinent manuellement le contenu doivent non seulement maitriser la définition du contenu terroriste, mais aussi connaître et entretenir leur connaissance des mots-clés et des expressions généralement utilisés par les organisations terroristes. Plus la personne qui évalue le contenu est familiarisée avec les pratiques terroristes (par exemple d'extrême droite, d'extrême gauche ou islamiste), plus il lui est facile de reconnaître leurs tactiques leur permettant de dissimuler leur caractère terroriste, comme l'utilisation de mots apparemment inoffensifs qui ont une signification idéologique codée au sein des milieux extrémistes[3]. Une base de données de logos et de symboles, c'est-à-dire d'éléments visuels indiquant un contexte terroriste, est également indispensable pour évaluer le contenu avec précision et rapidité.

[3] Åkerlund, M. (2022). Dog whistling extreme-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

Inclure les facteurs contextuels

La modération doit prendre en compte le contexte dans lequel le contenu est susceptible d'avoir été publié. Les facteurs contextuels pertinents comprennent, sans s'y limiter, (a) les conditions politiques, (b) les événements actuels, y compris les développements récents dans les actualités, et (c) les circonstances culturelles qui peuvent influencer les points de vue sur certaines questions. Il est souvent difficile d'en juger - en particulier lorsque les utilisateurs et utilisatrices sont anonymes et que le contenu qu'ils et elles publient contient peu ou pas d'indices textuels sur leur identité - mais dans certains cas, lorsque des facteurs externes sont mentionnés dans le contenu, les considérations contextuelles s'avèrent utiles.

En outre, lorsque vous avez affaire à un **contenu qui n'est pas terroriste mais qui est préjudiciable d'une autre manière** (par exemple, différentes formes d'incitation à la haine), vous pouvez également prendre en considération les points suivants.

Évaluer l'ampleur des dommages potentiels causés par le contenu

Les contenus terroristes, extrémistes et autres contenus offensants, en particulier lorsqu'ils incitent à commettre des infractions terroristes, peuvent causer des dommages importants. Lors de l'évaluation du contenu, il peut être utile de tenter de déterminer dans quelle mesure le contenu peut jouer un rôle dans ces dommages. Plus les dommages risquent d'être importants, plus l'action doit être rapide et bien réfléchie. Les effets psychologiques négatifs sur les personnes affectées par le contenu préjudiciable doivent être considérés comme une forme de dommage.

Tenir compte de l'impact potentiel de la suppression de contenu

Le règlement CTEL appelle à un équilibre délicat entre les droits fondamentaux pris en compte par la Charte (par exemple, la liberté d'expression et d'information) lors de l'examen des suppressions de contenu. Si le retrait n'a pas été ordonné par une autorité compétente, il existe d'autres moyens de traiter les contenus préjudiciables, dont vous trouverez des exemples dans le ▶ chapitre 3.

4. Que dois-je faire si je vois les choses différemment ? Comment contester une injonction de retrait reçue?

Que faire lorsque les évaluations du FSH et de l'autorité compétente se contredisent ?

Les fournisseurs de services d'hébergement et les fournisseurs de contenu ont le droit de contester les injonctions de retrait en vertu de l'article 9 du règlement CTEL. Ce droit est un important moyen de contrôle et d'équilibre qui garantit le respect des droits fondamentaux.

Les contestations des injonctions de retrait sont portées devant les tribunaux de l'État membre de l'UE dont l'autorité compétente a ordonné le retrait. Les **fournisseurs de services d'hébergement sont tenus de conserver dans des conditions sûres et pendant une période de six mois tout le contenu qui a été retiré (à la suite d'une injonction de retrait ou d'autres mesures), ainsi que toutes les données associées au contenu (par exemple, l'heure de la publication et les informations relatives au compte).**

Cela facilite les enquêtes et la prévention des infractions ou des menaces terroristes et permet de disposer d'un délai suffisant pour engager des actions en justice contre la suppression du contenu et, le cas échéant, pour rétablir le contenu (règlement CTEL, 27 & 28).

Cas particulier : Injonctions de retrait transfrontalières (règlement CTEL, article 4)

Si le FSH reçoit une injonction de retrait d'une autorité compétente d'un État membre de l'UE autre que son « autorité d'origine » (c'est-à-dire l'autorité compétente de l'État membre dans lequel le FSH a son principal établissement ou son représentant légal ; des détails sur les représentants légaux figurent dans le ▶chapitre 4), **une procédure spéciale s'applique.**

Lorsqu'elle établit une injonction de retrait, l'autorité compétente doit également en envoyer une copie à l'autorité d'origine, qui dispose d'un délai de 72 heures pour réexaminer la mesure. Si l'injonction de retrait porte atteinte aux droits et libertés fondamentaux consacrés par la Charte, elle peut prendre une décision motivée, qui peut constituer une objection.

Le fournisseur de services d'hébergement doit retirer et sécuriser le contenu dès réception de l'injonction de retrait (comme dans le cas d'une injonction de retrait prise par l'autorité d'origine). Le fournisseur de services d'hébergement (et la personne ayant fourni le contenu retiré) peut introduire une demande de réexamen auprès de l'autorité d'origine, qui est habilitée à réexaminer la mesure de retrait et à y répondre dans les 72 heures suivant la réception de la demande, après avoir notifié à l'autorité compétente qui a initialement ordonné le retrait qu'un réexamen est en cours.

Après que l'autorité de contrôle a rendu une décision motivée, les acteurs concernés (à savoir l'autorité d'origine, le fournisseur ou la fournisseuse de contenu, Europol le cas échéant) sont notifiés et, si l'injonction de retrait était une erreur, le contenu peut immédiatement être restauré.

Chapitre 3

Mettre en place de mécanismes de modération efficaces pour les contenus terroristes en ligne

Résumé : Contenu et points principaux de ce chapitre

- Dans le cadre de la modération de contenu, le contenu de l'utilisateur ou de l'utilisatrice **est examiné afin de déterminer si les règles légales** (par exemple, le règlement CTEL, le règlement sur les services numériques) **et les règles spécifiques à la plateforme** (par exemple, les conditions générales) ont **été respectées ou enfreintes.**
- **Si une règle a été violée, le contenu sera « modéré ».** Des mesures sont prises pour limiter la portée du contenu.
- Lorsqu'une autorité compétente **ordonne la suppression** d'un contenu sur la base du règlement CTEL, la **modération du contenu inclut toujours la suppression** du contenu terroriste concerné, bien que les fournisseurs de services d'hébergement et les utilisateurs/utilisatrices puissent contester cette décision s'ils et elles ne sont pas d'accord.
- Plusieurs points doivent être pris en compte avant, pendant ou après les mesures de modération prises par le FSH, ainsi que lors d'un examen ultérieur.
- Si les plateformes prennent des **mesures proactives contre d'autres formes de contenu potentiellement préjudiciables**, telles que les discours haineux ou les insultes, indépendamment du règlement CTEL, d'**autres approches de modération** peuvent également être envisagées.

1. Qu'est-ce que la modération de contenu et pourquoi est-elle nécessaire dans certains cas ?

La modération de contenu consiste à **examiner le contenu généré par l'utilisateur** (CGU) sur l'internet afin de pouvoir **évaluer le caractère approprié** du contenu sur la base des réglementations de la plateforme (par exemple les conditions générales) et des cadres juridiques (par exemple le règlement CTEL et le règlement sur les services numériques)[4]. Les **règlements de la plateforme et les exigences légales sont appliqués par le biais de la modération du contenu** afin que les mesures nécessaires soient prises à l'encontre des contenus interdits et/ou problématiques. Les fournisseurs de services d'hébergement sont tenus d'être

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

transparents sur les mesures spécifiques prises pour identifier et supprimer les contenus terroristes, y compris les processus proactifs et réactifs de modération des contenus et tout outil automatisé utilisé.

La modération de contenu est une procédure très importante, mais aussi très complexe et **sensible.** Disposer des ressources techniques et spécialisées appropriées peut représenter une charge financière importante pour les petites entreprises. En outre, le contenu terroriste peut être produit dans toutes les langues et il est extrêmement coûteux en ressources de s'assurer que les modérateurs sont en mesure d'évaluer le contenu dans ces conditions.

Les droits fondamentaux inscrits dans la Déclaration des Nations unies sur les droits de l'homme, en particulier le droit à la liberté d'expression, doivent toujours être mis en perspective avec le préjudice causé par le contenu en question. D'autre part, les fournisseurs de services d'hébergement ont la responsabilité sociale de limiter les contenus préjudiciables sur leur plateforme.

Une modération de contenu claire et transparente contribue à défendre les utilisateurs et les utilisatrice contre les effets d'un contenu préjudiciable et à instaurer la confiance. Du point de vue des fournisseurs de services d'hébergement, la modération des contenus peut permettre de limiter leur responsabilité, le respect des lois applicables et la protection contre les atteintes à la réputation causées par une utilisation abusive de la plateforme.

La modération du contenu ne signifie pas nécessairement sa suppression. Si la suppression d'un contenu peut être obligatoire en raison d'exigences légales - comme c'est le cas pour les injonctions de retrait basées sur le règlement CTEL - il existe d'autres moyens de traiter les contenus préjudiciables. En fonction de la gravité du contenu et des dommages qu'il peut causer, d'autres options peuvent être envisagées. Vous trouverez des exemples et des suggestions de stratégies de modération alternatives dans la ►section 3 de ce chapitre.

Les types de modération varient d'une plateforme à l'autre en fonction des fonctionnalités, des services offerts, des valeurs de la plateforme, ainsi que de la tolérance au risque et des différentes approches et stratégies[5]. Il est important d'être aussi transparent que possible en matière de modération de contenu. Cela inclut divers facteurs qui doivent être pris en compte avant, pendant et après la modération pour s'assurer que les utilisateurs et utilisatrices connaissent les actions de modération susceptibles d'être prises à l'encontre de leur contenu.

1. **Avant :** Avertir les utilisateurs et utilisatrices

Informer les utilisateurs et utilisatrices à un stade précoce des mesures de modération du contenu qui pourraient être prises à la suite d'un comportement ou d'un contenu interdit par les conditions générales ou par la loi. Cela implique d'inclure des informations sur vos stratégies de modération dans vos conditions d'utilisation (►chapitre 1). Cela permet également d'établir la confiance avec les utilisateurs et utilisatrices

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

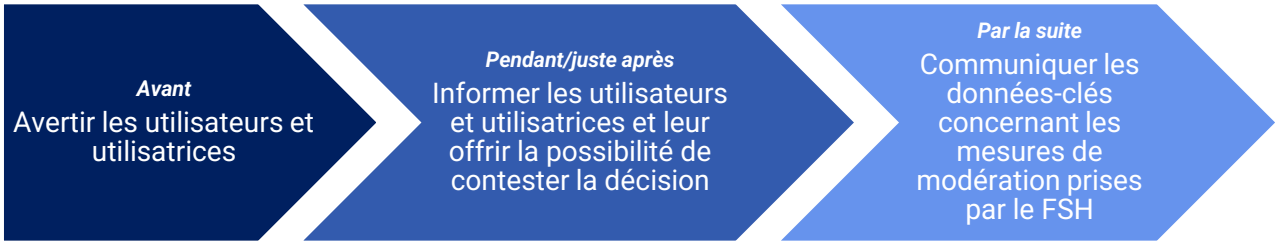
et de garantir sans ambiguïté que les contenus problématiques et illégaux ne seront pas tolérés sur la plateforme et que vous, en tant que FSH, voulez protéger les utilisateurs et utilisatrices contre de tels contenus.

2. **Pendant (ou juste après l'action) :** Informer les utilisateurs et utilisatrices dont le contenu a été bloqué ou modéré et leur offrir la possibilité de contester la décision.

L'article 11 du règlement CTEL oblige les fournisseurs de services d'hébergement à informer l'utilisateur ou l'utilisatrice qui a fourni (c'est-à-dire rédigé ou téléchargé) le contenu interdit (terroriste) qu'il a été bloqué par la plateforme, et à lui fournir, à sa demande, des informations sur le contexte ou l'injonction de retrait. L'obligation d'informer le fournisseur ou la fournisseuse de contenu ne s'applique pas, pour une pour une période n'excédant pas six semaines, si l'autorité compétente estime qu'il existe un danger particulier à divulguer la suppression. Lorsque la non-divulgation reste importante et appropriée, l'autorité compétente peut prolonger la période de six semaines supplémentaires (règlement CTEL, article 11). Le règlement CTEL (article 10) prévoit des mécanismes spécifiques pour gérer ces situations. Il est conseillé de mettre en place une procédure automatique standardisée par laquelle les fournisseurs et fournisseuses de contenu sont informés de la suppression et ont la possibilité de la contester. Ce mécanisme de notification et de contestation devrait également être utilisé dans les cas qui ne relèvent pas du règlement CTEL, mais qui ont émergé dans le cadre de mesures proactives de la part des modérateurs et modératrices de contenu des fournisseurs de services d'hébergement.

3. **Par la suite :** Communiquer régulièrement au public les données-clés concernant les mesures de modération du contenu prise par le FSH.

La transparence des décisions et des résultats de la modération des contenus est importante pour établir la confiance et la responsabilité entre un FSH et un utilisateur ou une utilisatrice, et est de plus en plus exigée par différents types de réglementations. Les rapports de transparence contiennent des informations sur les décisions de modération de la plateforme, concernant non seulement le nombre et le type de violations identifiées, mais aussi la manière exacte dont ces violations ont été traitées. (Vous trouverez plus de détails sur les rapports de transparence et les exigences de transparence du CTEL dans le ▶ chapitre 6).



2. Conseils pratiques : Comment mettre en œuvre la modération de contenu ?

Il est important que les fournisseurs de services d'hébergement disposent d'un **processus** clair **de modération des contenus** qui puisse être **adapté** à leurs services et à leurs besoins commerciaux spécifiques. Il est essentiel de commencer par identifier les contenus interdits et être en mesure d'évaluer leur illégalité, comme indiqué plus en détail dans le ▶chapitre 2. Pour rappel, voici les étapes-clés :

1. Définition du contenu terroriste
2. Optimisation grâce aux outils automatisés
3. Aide des systèmes de signalement
4. Examen par des modérateurs et modératrices humains
5. **Décision sur la manière de traiter le contenu (c'est-à-dire décision sur la modération du contenu)**
6. Notification du fournisseur/de la fournisseuse de contenu

Dans cette section, nous nous intéressons à l'*étape 5*. Dans le ▶chapitre 2, nous avons nommé cette étape « Décision sur la manière de traiter le contenu ». Elle peut être considérée comme synonyme de « décision sur la modération du contenu ». Si vous avez reçu au préalable une injonction de retrait de la part d'une autorité compétente, vous devez retirer le contenu ou contester l'injonction. Les fournisseurs de services d'hébergement peuvent également décider d'identifier et de supprimer les contenus terroristes ou interdits de manière proactive.

Ce guide distingue cinq types de modération : pré-, post-, réactive, distribuée et automatisée. De brèves explications ainsi que leurs avantages et inconvénients sont présentés ci-après.

Types de modération	Explications
Pré-modération	Les modérateurs et modératrices vérifient le contenu avant qu'il ne soit publié. **Avantages :** Haut degré de sécurité et de conformité du contenu avec les normes juridiques et les réglementations propres à la plateforme. **Inconvénients :** charge importante en termes d'efforts, de personnel et (éventuellement) de coûts financiers.
Post-modération	Les modérateurs examinent le contenu immédiatement après la publication. **Avantages :** Permet des interactions rapides entre l'utilisateur/l'utilisatrice et le contenu (en raison de la publication immédiate). **Inconvénients :** Beaucoup de temps, de personnel (et éventuellement de coûts financiers) ; dans le cas de contenus interdits ou préjudiciables, les utilisateurs et utilisatrices sont exposés à ces contenus.
Modération réactive	Les modérateurs et modératrices examinent le contenu après que les utilisateurs et utilisatrices l'ont signalé.

	Avantages : Moins d'efforts de la part du FSH ; renforcement de la confiance des utilisateurs et utilisatrices (grâce à la possibilité d'établir des rapports). **Inconvénients :** Responsabilité des utilisateurs et utilisatrices ; en cas de contenu interdit/préjudiciable, les utilisateurs et utilisatrices sont toujours exposés au contenu ; risque de fausses alertes nécessitant des efforts supplémentaires de la part du fournisseur de services d'hébergement.
Modération distribuée	La communauté d'utilisateurs et d'utilisatrices joue le rôle de modérateur, souvent par le biais d'un mécanisme de vote positif et négatif qui mesure la fiabilité du contenu et est finalement utilisé pour déterminer la portée publique du contenu (c'est-à-dire qu'un contenu ayant beaucoup de votes positifs est mieux classé et a une plus grande portée). **Pour :** Moins de ressources de la part du FSH ; encouragement de l'engagement des utilisateurs et utilisatrices ; autorégulation. **Inconvénients :** Responsabilité des utilisateurs et utilisatrices ; en cas de contenu interdit/préjudiciable, les utilisateurs et utilisatrices sont exposés au contenu ; risque de comportement manipulateur coordonné de la part d'utilisateurs et utilisatrices malveillants.
Modération automatisée	Les modèles basés sur l'intelligence artificielle (par exemple, les filtres, les algorithmes) jouent le rôle de modérateur. **Avantages :** Moins gourmand en ressources une fois mis en place ; détection précoce, rapide et hautement évolutive des contenus malveillants. **Inconvénients :** Risque d'erreurs, notamment en ce qui concerne la suppression erronée de contenu (problématique dans le contexte de la liberté d'expression, d'où l'importance de l'évaluation humaine) ; nécessité d'une maintenance continue et d'une adaptation aux nouveaux développements.

(D'après Grimes-Viort, 2010)[6]

Il est recommandé de combiner plusieurs approches de modération de contenu. La modération réactive, par exemple, peut être facilement combinée avec la pré-modération ou la post-modération. Dans la plupart des cas, les **combinaisons sont très utiles, et dans certains cas, elles peuvent même être nécessaires**. Les décisions de modération automatisée doivent toujours faire l'objet d'un examen humain afin d'éviter les violations des droits fondamentaux des utilisateurs et utilisatrices et les risques pour la liberté d'expression.

Les acteurs et actrices extrémistes et terroristes peuvent utiliser des **tactiques pour éviter les méthodes de modération de contenu connues**, opérer sous le radar et ainsi contourner la modération, en particulier les mécanismes de modération de contenu automatisés. Parmi les tactiques les plus courantes, on peut citer le *raccourcissement d'URL* pour éviter les filtres ou les blocages de sites web, la création de comptes et de

[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

contenus en *miroir*, qui consiste à publier ou à créer plusieurs fois des contenus/comptes identiques afin de submerger les modérateurs/modératrices et de disposer de copies de sauvegarde en cas de suppression, ou encore le *choix délibéré d'une orthographe (incorrecte) d'un mot* afin de tromper les systèmes de détection automatiques. L'éventail des techniques d'évitement souligne l'importance de la modération humaine.

Votre lecture de ce guide, ou votre participation au cours en ligne (qui donne lieu à un certificat inter-universitaire) qui est également proposé dans le cadre du projet TATE, montre que vous vous intéressez à ces problématiques. Vous trouverez d'autres exemples de contournement de la modération de contenu ainsi que des réponses possibles et efficaces dans la plateforme de partage des connaissances de Tech Against Terrorism, disponible en anglais ici.

3. Autres approches de modération

La modération ne doit pas toujours être synonyme de suppression de contenu. D'autres approches, telles que celles suggérées ci-après, peuvent intéresser les fournisseurs de services d'hébergement qui cherchent à modérer le contenu (pro-)activement au-delà des exigences du règlement CTEL. Il est important de souligner que ces approches de modération alternatives sont **hors du champ d'application du règlement CTEL et ne peuvent être utilisées que lorsque les plateformes n'ont pas reçu d'injonction de retrait mais souhaitent modérer de manière proactive des contenus non terroristes et autrement préjudiciables.** Lorsque vous recevez un injonction de retrait, la procédure est claire : vous devez retirer le contenu et aucune approche de modération alternative n'entre en ligne de compte.

VOUS VOULEZ EN FAIRE PLUS POUR ASSURER LA SÉCURITÉ DE VOTRE PLATE-FORME ?

Masquer le contenu

Les fournisseurs de services d'hébergement peuvent masquer partiellement ou totalement un contenu, et ainsi éviter de le bloquer, s'ils estiment que les utilisateurs et utilisatrices peuvent trouver le contenu offensant ou répréhensible, mais qu'il est néanmoins légitime, légal et autorisé sur l'ensemble de la plateforme. Le fait de cacher un contenu aux personnes appartenant à un groupe vulnérable ou aux personnes situées dans un pays où le contenu est illégal (alors qu'il est autorisé dans d'autres pays) est une alternative à la suppression. Diverses fonctionnalités techniques peuvent être utilisées pour masquer le contenu, telles que des filtres de connexion ou de paywall, un mode de recherche sécurisé pour afficher un contenu adapté à l'âge et un blocage géographique ou temporel.

Désactiver les métriques de récompense (« désengagement »)

Le désengagement prive certains contenus ou utilisateurs/utilisatrices de certaines métriques d'engagement, décourage l'activité autour de la publication et peut rendre le contenu généralement peu gratifiant à publier. Toutefois, le contenu et le compte d'utilisateur restent sur la plateforme. Le désengagement limite l'importance de ces publications ou des comptes sur la plateforme. Les tactiques de désengagement

habituelles comprennent la désactivation des fonctions de la plateforme, telles que, dans le cas de nombreux réseaux sociaux, la possibilité d'aimer, de commenter ou de partager des publications, de sorte qu'elles peuvent uniquement être lues, la démonétisation (c'est-à-dire priver les comptes de la possibilité de gagner de l'argent grâce à leur contenu) ou la dé-vérification (c'est-à-dire supprimer toute certification de l'identité du compte ou de l'utilisateur/utilisatrice). Ces sanctions peuvent entraîner un changement dans le traitement du contenu ou du compte par l'algorithme de la plateforme : le déclassement du contenu signifie qu'il est plus difficile de le diffuser à grande échelle et de le promouvoir par le biais des mécanismes de la plateforme (généralement des algorithmes de recommandation).

Ajouter des notes pédagogiques au contenu

L'objectif de l'approche pédagogique est d'offrir aux utilisateurs et utilisatrices des informations supplémentaires afin qu'ils et elles puissent décider par eux-mêmes et elles-mêmes s'ils et elles veulent voir le contenu ou non. En fin de compte, c'est la plateforme qui décide quel contenu est accompagné de telles notes, ce que ces notes impliquent, de quelle catégorie de préjudice les utilisateurs et utilisatrices sont avertis et quelle quantité d'informations supplémentaires est proposée. Une pratique bien connue de Twitter consiste à avertir les utilisateurs qu'un tweet peut contenir un contenu préjudiciable, tel que des informations erronées ou des théories complotistes, et à ne permettre aux utilisateurs et utilisatrices de voir le contenu qu'après avoir activement confirmé qu'ils et elles le souhaitent, en cliquant sur un bouton. En particulier dans le cas de contenus politico-idéologiques susceptibles de favoriser la radicalisation, des contre-récits et des liens vers des informations éducatives peuvent sensibiliser les gens aux effets possibles de ces contenus.

Responsabiliser les utilisateurs et utilisatrices (« responsabilisation de la communauté »)

Les mécanismes de modération fondés sur l'autonomisation des communautés ont pour principe de permettre aux utilisateurs et utilisatrices de créer l'espace numérique qu'ils et elles imaginent. Ces stratégies peuvent présenter un intérêt particulier pour les plateformes où l'idée de communauté est importante, ou dont les pratiques de modération dépendent déjà, dans une certaine mesure, du soutien des utilisateurs et utilisatrices. Ces approches de modération relèvent de la modération distribuée. Outre la fonctionnalité de vote positif et négatif déjà mentionnée, il s'agit également du blocage individuel, qu'un grand nombre de plateformes proposent déjà, ou de l'utilisation d'administrateurs ou de modérateurs, hommes et femmes, issus de la communauté elle-même. Le concept de « signaleurs de confiance », pris en compte dans le règlement sur les services numériques (article 22), est étroitement lié à cette question. Ce concept fait référence aux utilisateurs et utilisatrices qui sont particulièrement dignes de confiance et compétents pour évaluer l'illégalité d'un contenu et le signaler (objectivement et rapidement) et qui représentent des intérêts collectifs (orientés vers le bien-être public) indépendamment de la plateforme. Les

> contenus signalés de cette manière doivent être traités en priorité et rapidement par le FSH.

Ces approches alternatives de la modération peuvent être pertinentes même lorsque les plateformes ne sont pas obligées par le règlement CTEL de prendre des mesures. Le règlement CTEL promeut une approche proactive : si, dans le cadre de mesures de modération proactives, le FSH détecte du contenu en lien avec une menace imminente pour la vie ou un acte terroriste, celui-ci doit être supprimé et l'autorité compétente de l'État membre de l'UE concerné doit être informée immédiatement (règlement CTEL, article 14.5).

Tech Against Terrorism fournit ici plus de détails sur les techniques requises par ces approches alternatives, sur leurs avantages et inconvénients et sur des études de cas.

Chapitre 4

Établir des points de contact et des représentants légaux

Résumé : Contenu et points principaux de ce chapitre

- Le règlement CTEL établit une distinction entre les points de contact et les représentants légaux.
- **Des points de contact doivent être mis en place par chaque FSH** et sont **chargés de** recevoir les **injonctions de retrait** et de les traiter rapidement.
- **Si les fournisseurs de services d'hébergement ne sont pas établis dans l'UE, un représentant légal doit également être désigné**. Cette personne est responsable de la réception, du respect et de l'application du règlement CTEL. Le représentant légal peut, mais ne doit pas, être également le point de contact.
- Chaque FSH, qu'il ait été exposé ou non à des contenus terroristes, doit établir un point de contact et, le cas échéant, un représentant légal, conformément au règlement CTEL.

Le règlement CTEL stipule que *tous les* FSH qui sont concernés par ce règlement (c'est-à-dire qui tombent dans le champ d'application du règlement par définition ; ►la section sur les plateformes concernées dans l'introduction) sont obligés de nommer un point de contact ou un représentant légal.

1. Que sont les points de contact et les représentants légaux ?

Point de contact (règlement CTEL, (42) & article 15)

- **Objectif** : Les points de contact du FSH facilitent le traitement immédiat des injonctions de retrait. Le point de contact a donc une fonction opérationnelle.
- **Logistique** : Le point de contact doit être en mesure de recevoir et de transmettre les injonctions de retrait par voie électronique, qu'il soit en interne ou externalisé.
- **Ressources nécessaires** : Le point de contact doit disposer d'une capacité technique et d'un accès suffisants, et être doté d'un personnel suffisant pour que les injonctions de retrait puissent être traités sans délai. Étant donné que les contenus terroristes doivent être retirés dans l'heure qui suit la réception de l'injonction de retrait, le point de contact doit être disponible 24 heures sur 24 et 7 jours sur 7.
- **Localisation** : Le point de contact ne doit pas nécessairement être situé dans l'UE.
- **Communication** : La langue dans laquelle communiquer avec le point de contact doit être indiquée dans les informations de contact. Afin de permettre la communication entre les FSH et les autorités compétentes des États membres, il convient d'utiliser au moins une langue officielle de l'UE. Il doit s'agir d'une langue dans laquelle les conditions générales de la plateforme sont également disponibles.

Représentant légal (règlement CTEL, article 17)

- **Nécessité** : Si le FSH n'a pas son établissement principal dans l'UE, un représentant légal doit être désigné. Il s'agit d'une personne physique ou morale située dans l'un des États membres de l'UE dans lequel le FSH offre ses services.
- **Objet** : Ce représentant légal est chargé de recevoir, de respecter et d'appliquer le règlement CTEL et les injonctions de retrait en particulier.
- **Ressources** : Les fournisseurs de services d'hébergement doivent fournir aux représentants légaux le pouvoir, les capacités et les ressources nécessaires pour se conformer au règlement CTEL et coopérer avec les autorités compétentes.
- **Responsabilité** : Les représentants légaux peuvent être tenus pour responsables des violations du règlement CTEL.
- **Relation avec le point de contact** : Le représentant légal peut également faire office de point de contact mais ce n'est pas obligatoire.

2. Pourquoi est-il nécessaire d'avoir un point de contact ou un représentant légal ?

Les autres chapitres abordent de nombreuses options qui peuvent être intéressantes pour les fournisseurs de services d'hébergement d'un point de vue commercial, mais le principal argument en faveur de la mise en place d'**un point de contact ou de la désignation d'un représentant légal** est le simple fait que c'est obligatoire. Cela peut impliquer une certaine réaffectation des ressources, mais il est incontestablement dans l'intérêt des fournisseurs de services d'hébergement de se conformer à la loi et d'éviter les conséquences négatives de la non-conformité, telles que la perte de réputation, de confiance et d'argent, si des amendes sont imposées.

Les fournisseurs de services d'hébergement doivent permettre aux autorités compétentes de consulter les informations relatives au point de contact et, par la suite, de procéder à des notifications électroniques (c'est-à-dire de soumettre une injonction de retrait par voie électronique). En règle générale, cela implique de fournir une adresse électronique (par exemple, dans la section « Contactez-nous » du site web du fournisseur de services d'hébergement) par laquelle l'autorité compétente peut contacter le point de contact.

Deux aspects de la désignation d'un représentant légal doivent être soulignés. Premièrement, l'**identité du représentant légal doit être rendue publique** (voir point de contact). Deuxièmement, **le FSH doit informer l'autorité d'origine** (c'est-à-dire l'autorité compétente de l'État membre où le représentant légal est établi).

Europol a également développé une plateforme, la Plateforme Européenne de Retraits de Contenus illicites sur Internet (**PERCI**)**, pour soutenir la mise en œuvre du règlement CTEL**. L'objectif du PERCI est de garantir que les fournisseurs de services d'hébergement puissent recevoir les ordres de retrait des États membres par le biais d'un canal sécurisé commun, au lieu de 27 systèmes distincts par État membre. Il rationalise les injonctions de retrait provenant de différents États membres et sert de point de contact unique dans le cadre des signalements et des injonctions de retrait, et vise à éviter la duplication des injonctions de retrait, c'est-à-dire l'envoi d'une même mesure par deux États membres. Il est également utile pour les FSH de disposer d'une réception centralisée des injonctions de retrait et des signalements reçus au fil du temps, afin de

communiquer à leur sujet de façon transparente. Grâce au PERCI, les FSH peuvent également demander un examen et un réexamen pour contester une injonction de retrait.

3. Qu'est-ce que l'autorité compétente d'un État membre de l'UE et comment la contacter ?

Dans certaines circonstances, il est nécessaire de contacter l'autorité compétente au sujet du représentant légal. L'une de ces circonstances est la contestation d'une injonction de retrait. Un autre cas de figure serait, par exemple, que le fournisseur de services d'hébergement prenne connaissance d'un contenu terroriste impliquant une menace imminente pour la vie ou un acte terroriste, en dehors du contexte d'une injonction de retrait. Dans ce cas, le fournisseur de services d'hébergement est tenu de le supprimer immédiatement et d'en informer l'autorité compétente de l'État membre concerné (règlement CTEL, article 14.5).

La plupart des autorités compétentes des États membres de l'UE ont déjà mis en place des points de contact. Une liste actualisée comprenant les coordonnées des personnes à contacter est disponible sur le site web de la Commission européenne ici.

Chapitre 5

Mettre en place d'un système de notification et de réclamation concernant les contenus supprimés

Résumé : Contenu et points principaux de ce chapitre

- **Les utilisateurs et utilisatrices** concernés **peuvent utiliser les procédures de plainte pour contester les mesures de suppression** (et, si nécessaire et en dehors du champ d'application du CTEL, d'autres mesures de modération proactives).
- Les procédures de réclamation sont importantes en tant que **mécanisme de contrôle et de feedback**, du point de vue des utilisateurs, des utilisatrices, des entreprises et de la loi.
- Tout fournisseur de services d'hébergement « établit un mécanisme efficace et accessible » (règlement CTEL, article 10.1).
- Selon le règlement CTEL, les systèmes de réclamation doivent répondre à certaines exigences techniques et de contenu.
- Une procédure de plainte peut aboutir à deux résultats différents, à savoir (1) la plainte est acceptée parce qu'il est établi que le contenu a été bloqué à tort, ou (2) la plainte est rejetée parce qu'il est établi que le contenu a été bloqué à juste titre.
- En fonction de l'issue de la procédure de plainte, le contenu peut faire l'objet d'autres mesures.
- Ce chapitre fournit également des conseils sur la conception et la mise en œuvre d'une procédure de plainte adaptée à chaque FSH.

Les procédures de plainte permettent aux utilisateurs et utilisatrices de contester les suppressions de contenu en communiquant avec la plateforme et **constituent la première étape pour contester (légalement) une injonction de retrait.** Pour en savoir plus sur la procédure de contestation, consultez le ▶chapitre 2.

1. Pourquoi est-il nécessaire de mettre en place un mécanisme de réclamation transparent ?

Il est important de mettre en place un mécanisme de réclamation, en particulier (a) d'un point de vue juridique, (b) du point de vue de l'utilisateur/utilisatrice et (c) du point de vue de l'entreprise.

a) Point de vue juridique

Les procédures de plainte sont conformes aux **réglementations établies par la loi** qui **rendent ces mécanismes nécessaires**. Par exemple, le règlement CTEL (règlement CTEL, article 10) stipule que les fournisseurs de services d'hébergement doivent mettre en place un mécanisme de réclamation efficace et

accessible pour donner aux utilisateurs et utilisatrices la possibilité de contester la suppression ou le blocage du contenu après une mesure spécifique. En vertu du règlement CTEL, la personne ayant fourni le contenu doit être informé du résultat final dans un délai de deux semaines.

b) Point de vue de l'utilisateur/utilisatrice

Un mécanisme de plainte n'est pas simplement une obligation légale. Un mécanisme de plainte clair et accessible contribue à instaurer la confiance des utilisateurs et utilisatrices. De cette manière, les fournisseurs de services d'hébergement mettent en évidence que les utilisateurs et utilisatrices peuvent avoir confiance dans les processus de modération, que ces processus sont fondés sur un sentiment de **responsabilité envers les utilisateurs et utilisatrices** et que les plateformes tiennent compte des **droits fondamentaux** tels que la liberté d'expression et d'information.

c) Point de vue de l'entreprise

Les mécanismes de plainte peuvent constituer une forme utile d'**autocontrôle**, qui permet d'évaluer l'efficacité, la justesse et la cohérence des mesures et des normes de modération. Ces mécanismes peuvent donner l'assurance que votre plateforme est utilisée en conformité avec les conditions générales, ce qui contribue à protéger la **réputation** des services en ligne et le droit à la liberté d'expression en ligne.

2. Quelles sont les exigences d'un système de réclamation ?

Selon le règlement CTEL (33 & article 10), les systèmes de réclamation doivent :

- Être faciles d'utilisation,
- Être efficaces et (facilement) accessibles,
- Fournir un cadre sûr dans lequel les plaintes sont traitées rapidement et de manière transparente, de sorte que le plaignant ou la plaignante soit informé du résultat de l'examen dans un délai de deux semaines.

Dans le cadre du règlement CTEL, les systèmes de plainte doivent être mis en place pour pouvoir restaurer les contenus retirés par erreur. Toutefois, des plaintes peuvent également être déposées contre des mesures prises pour faire respecter les conditions générales de la plateforme en dehors du champ d'application du règlement CTEL.

3. Comment les plaintes sont-elles être traitées et quels sont les résultats possibles ?

Une fois que le système de plainte est mis en œuvre et que les utilisateurs et utilisatrices ont déposé une plainte, le FSH **examine la plainte** et **communique la décision au plaignant ou à la plaignante dans un délai maximum de deux semaines pour les cas relevant du règlement CTEL**.

Le processus d'examen peut déboucher sur deux résultats possibles, énumérés dans le tableau suivant.

Résultat A **La plainte d'un utilisateur/d'une utilisatrice contre la suppression d'un contenu est acceptée**	*Résultat B* **La plainte d'un utilisateur/d'une utilisatrice contre la suppression de contenu est rejetée**
Résultat : La plainte contre la suppression du contenu est **justifiée** et est donc acceptée.	**Résultat :** La plainte contre la suppression du contenu n'est **pas justifiée** et est donc rejetée.
Signification : Le contenu a été retiré, supprimé ou modéré de manière inappropriée.	**Signification :** Le contenu a été correctement retiré, supprimé ou modéré.
Procédure ultérieure : Le fournisseur de services d'hébergement (1) informe le plaignant ou la plaignante du résultat de l'examen et (2) rétablit le contenu.	**Suite de la procédure :** Le FSH (1) informe le plaignant ou la plaignante du résultat de l'examen et (2) lui donne les raisons de cette décision.

4. Conseils pratiques : Quels sont les éléments utiles pour mettre en place un système de réclamation ?

Le fait qu'un système de réclamation doive présenter certaines caractéristiques n'est pas seulement une exigence de la loi, mais aussi une exigence de **convivialité**.

Comme d'autres questions-clés, telles que les conditions générales (▶chapitre1), la mise en place d'un processus d'identification des contenus interdits (▶chapitre 2), ou le choix de mécanismes de modération spécifiques à la plateforme (▶chapitre 3), la méthode de soumission d'une plainte peut varier considérablement d'une plateforme à l'autre. Alors que certaines plateformes peuvent proposer de contester une décision par courrier électronique, d'autres choisissent de mettre en œuvre des applications standardisées basées sur des formulaires. Les **systèmes de réclamation doivent être adaptés** à l'objectif, à la structure et à l'organisation de la plateforme. La check-list suivante fournit des conseils sur la mise en place d'un système de réclamation.

Donner des informations claires sur le retrait du contenu

Notifiez la personne dont le contenu a été supprimé. Envisagez de le faire même si le contenu a été modéré d'une autre manière. Dans ce contexte, informez également la personne des raisons pour lesquelles le contenu a été retiré (voir le point suivant sur les informations relatives au contexte éducatif et pédagogique) et de la manière dont elle peut contester la décision de retrait (voir le point suivant sur l'explication de la procédure de plainte).

Expliquer la procédure de plainte

Expliquez à vos utilisateurs et utilisatrices la procédure de réclamation. Cela doit se faire lors de la notification de la suppression ou de la modération d'un contenu. Des informations sur la procédure de réclamation peuvent également être incluses dans les conditions générales. L'explication de la procédure de réclamation doit comprendre (1) comment les plaintes peuvent être déposées, (2) comment fonctionne la procédure d'examen et (3) comment les utilisateurs et utilisatrices sont informés du résultat de l'examen.

Fournir des informations explicatives

Offrir aux utilisateurs des informations afin d'expliquer pourquoi le contenu a été retiré et quelles sont les conditions qu'il enfreint. Par exemple, le contenu peut être retiré s'il viole les conditions générales et/ou si une injonction de retrait a été envoyée par une autorité compétente dans le cadre du règlement CTEL. Dans le premier cas, cette information peut être précisée en ajoutant les interdictions spécifiques qui ont été violées (par exemple, contre les discours haineux, l'incitation à la violence, le contenu sexuel et le harcèlement). Dans le second cas, il est recommandé d'ajouter au système de plainte un bref résumé du règlement CTEL, afin que le cadre juridique soit clair pour l'utilisateur ou l'utilisatrice.

Fournir des informations régulières sur l'état d'avancement de la plainte

Les utilisateurs et utilisatrices doivent recevoir des mises à jour régulières sur l'avancement de la plainte afin de leur montrer que la procédure se poursuit. Au minimum, ces mises à jour doivent consister dans une notification lorsque l'examen est terminé et dans une communication du résultat. Une communication plus détaillée pourrait informer l'utilisateur ou l'utilisatrice que la plainte a été reçue et qu'elle est en cours d'examen par le personnel du FSH. La meilleure pratique consisterait à indiquer un délai pour l'examen de sa plainte. Ces informations peuvent être communiquées par courrier électronique ou affichées en ligne.

Documenter la procédure de plainte (individuelle)

Il est important de documenter la procédure de plainte. Ce document servira de référence en cas de questions ou de litiges ultérieurs.

Chapitre 6

Aide et conseils pratiques concernant les rapports de transparence

Résumé : Contenu et points principaux de ce chapitre

- Les rapports de transparence permettent aux fournisseurs de services d'hébergement de **communiquer publiquement** sur la manière dont les **valeurs** de la plateforme sont respectées, ainsi que les **mesures prises à l'encontre des contenus et des comportements interdits et illégaux.**
- Les rapports de transparence sont un **outil important pour assumer la responsabilité publique et démontrer la crédibilité et la fiabilité du FSH.**
- Différentes **lois**, dont le règlement CTEL, **exigent** explicitement **des rapports de transparence**.
- Lors de la préparation des rapports de transparence, il est conseillé de procéder de façon systématique, avant, pendant et après le processus, en suivant des étapes spécifiques.
- Le **règlement CTEL impose aux fournisseurs de services d'hébergement qui sont exposés à des contenus terroristes ou qui ont pris des mesures à leur encontre de publier un rapport annuel de transparence sur leurs activités relatives aux contenus terroristes**. Ce rapport doit contenir certains chiffres et informations essentiels, tels que le nombre de demandes de retrait reçues, le nombre de contenus supprimés et le nombre d'utilisateurs et utilisatrices interdits. Le rapport doit être publié au plus tard le 1er mars de l'année suivante (règlement CTEL, article 7.2).

1. Qu'est-ce qu'un rapport de transparence ?

Les rapports de transparence sont un **outil important permettant aux FSH de prouver leur responsabilité, leur crédibilité et leur fiabilité**. Les rapports de transparence **contiennent des données importantes sur les demandes que les FSH reçoivent de la part des autorités compétentes dans le monde entier et sur la manière dont ces demandes sont traitées**, ce qui vise à rendre transparentes la collaboration et la coopération avec les autorités et d'autres organes étatiques[7].

Les rapports de transparence donnent également un aperçu des mesures prises par le FSH pour faire respecter la réglementation (par exemple, par le retrait de contenu ou d'autres mesures de modération). Cela comprend l'application (1) des politiques spécifiques à la plateforme (généralement les conditions

[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

générales ; ▶chapitre 1), (2) des droits tels que le droit d'auteur ou le droit des marques, et (3) de la législation et de la réglementation (locales) qui entraînent le retrait de contenu[8] . Dans l'UE, la législation locale comprend le règlement CTEL et la loi sur les services numériques. La législation spécifique à un pays peut également être pertinente, comme la loi sur l'application de la loi sur les réseaux en Allemagne.

Les rapports de transparence sont généralement publiés régulièrement. Le règlement CTEL stipule que cela doit être fait **(au moins) une fois par an** si un fournisseur de services d'hébergement a pris des mesures contre du contenu terroriste (règlement CTEL, 30).

D'un fournisseur de services d'hébergement à l'autre, les rapports de transparence et les indicateurs qui y figurent peuvent varier considérablement[9]. **Pour les rapports de transparence conformes au règlement CTEL, il existe des exigences spécifiques concernant les informations à y inclure.** Vous trouverez plus d'informations à ce sujet dans la ▶section 4 de ce chapitre.

2. Pourquoi les rapports de transparence sont-ils nécessaires ?

Les rapports de transparence **permettent aux utilisateurs et utilisatrices et aux tiers d'évaluer dans quelle mesure les fournisseurs de services d'hébergement restent fidèles à leurs valeurs, aux exigences légales et à la protection des données et de la vie privée**[6]. Examinons les différents points de vue sur l'importance des rapports de transparence.

a) Point de vue juridique

Les rapports de transparence sont utiles et souvent nécessaires pour se conformer aux règlements de l'UE tels que le CTEL, le règlement sur les services numériques et, le cas échéant, les lois spécifiques à chaque pays. Si les fournisseurs de services d'hébergement ont pris des mesures contre la diffusion de contenus terroristes au cours d'une année civile, que ce soit de manière proactive ou en application d'une mesure de retrait, un rapport de transparence **doit être publié au** plus tard le **1er mars de l'année suivante** (règlement CTEL, article 7.2). Les autorités compétentes sont également tenues de publier des rapports de transparence annuels (règlement CTEL, 31).

b) Point de vue des utilisateurs/utilisatrices et des tiers

Les rapports de transparence aident les utilisateurs/utilisatrices et les tiers à évaluer dans quelle mesure les FSH s'acquittent de leur **responsabilité vis-à-vis de la société**. Dans le même temps, la publication régulière des rapports de transparence aide le FSH à gagner la **confiance du public et à se forger une bonne réputation** en démontrant sa conformité, son engagement et sa fiabilité.

[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/

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

c) Point de vue de l'entreprise

À l'instar des mécanismes de plainte, les rapports de transparence peuvent également constituer un **outil d'autocontrôle** permettant d'identifier les domaines dans lesquels les processus opérationnels peuvent être optimisés. Étant donné que les petits fournisseurs de services d'hébergement sont particulièrement populaires auprès des acteurs terroristes (pour plus de détails, voir le rapport Tech Against Terrorism), il leur incombe de prendre des mesures pour lutter contre la diffusion de contenus terroristes. Les rapports de transparence sont un moyen pour les petits fournisseurs de services d'hébergement de démontrer leur engagement dans ce domaine.

3. Un processus de préparation des rapports de transparence

La mise en place initiale et la remise d'un rapport de transparence peuvent être une tâche fastidieuse. Cependant, **une fois qu'un processus et une routine de préparation des rapports de transparence annuels ont été établis**, il n'est généralement pas nécessaire de retravailler complètement la structure du rapport, et des mises à jour sont généralement suffisantes. Ce guide explique comment établir un rapport de transparence du point de vue du fournisseur de services d'hébergement, ainsi que les éléments à prendre en compte avant, pendant et après le processus de création.

a) Avant l'élaboration du rapport de transparence

Analyser le paysage juridique

Obtenez une vue d'ensemble des réglementations légales qui s'appliquent à votre fournisseur de services d'hébergement. Par exemple, le règlement CTEL et le règlement sur les services numériques s'appliquent à l'ensemble de l'UE. Il peut y avoir des réglementations spécifiques à un pays ainsi que des exigences concernant d'autres sujets pertinents pour votre fournisseur de services d'hébergement, en dehors du contenu terroriste. Si vous travaillez avec une équipe juridique, il est souvent conseillé d'échanger avec eux afin de déterminer quelles réglementations (en dehors du règlement CTEL dont il est question ici) s'appliquent à votre fournisseur de services d'hébergement.

Déterminer les objectifs du rapport de transparence

Réfléchissez aux objectifs de votre rapport de transparence. Les questions-clés qui peuvent vous aider sont les suivantes : Souhaitez-vous simplement vous acquitter de vos obligations légales ou souhaitez-vous également aborder d'autres aspects et détailler votre engagement les concernant ? Qui voulez-vous atteindre, c'est-à-dire quel groupe cible est pertinent pour votre FSH (par exemple, les acteurs et actrices politiques, les utilisateurs et utilisatrices, les financiers) ? À quelle fréquence souhaitez-vous publier des rapports de transparence et quel est le meilleur moment pour les communiquer ?

Déterminer quelles données peuvent et doivent être incluses

Déterminez les données que vous allez inclure dans le rapport de transparence. D'une part, la *capacité* concrète est à prendre en compte, c'est-à-dire quelles données sont concrètement disponibles ? D'autre part, quelles sont les exigences légales que vous devez respecter et quelles sont les données

nécessaires à cet effet ? Vous trouverez dans le ▶sous-chapitre suivant les informations et les données que vous devez inclure conformément au règlement CTEL.

b) Lors de l'élaboration du rapport de transparence

Utiliser un langage clair et concis

Lors de la création du rapport de transparence, veillez à utiliser un niveau de langue clair et concis. Cela facilite l'engagement et la compréhension d'un matériel complexe. Le langage doit être adapté au lectorat envisagé.

Fournir des informations contextuelles et des explications

Fournissez à vos lecteurs et lectrices des informations et des explications contextuelles. Ces explications leur permettront de mieux comprendre le mode de fonctionnement de votre FSH et les raisons pour lesquelles vous fournissez éventuellement un rapport de transparence plus détaillé qu'un simple rapport répondant uniquement aux exigences réglementaires minimales.

Intégrer le feedback interne

Lors de la programmation d'un rapport de transparence, il convient d'intégrer les feedbacks internes, qui peuvent porter sur le contenu et/ou sur la langue.

c) Après l'élaboration du rapport de transparence

Publier le rapport de transparence

Pensez aux langues dans lesquelles vous souhaitez publier le rapport de transparence et créez les traductions appropriées. Déterminez également où le rapport doit être accessible, c'est-à-dire où sur votre site web. En outre, vous pouvez également envisager d'inclure le rapport de transparence dans divers supports de communication afin d'en optimiser la visibilité. Il peut s'agir de l'intégrer à votre site web, de l'envoyer dans des bulletins d'information par courriel ou de le partager sur les médias sociaux.

Mettre régulièrement à jour les données et le rapport de transparence

Lorsque le premier rapport de transparence est terminé, vous avez déjà jeté les bases qui faciliteront le suivant, puisque vous avez déjà créé un modèle. Le processus de collecte des données pertinentes tout au long de l'année et leur organisation pour les rendre facilement accessibles faciliteront la production du prochain rapport de transparence et réduiront le temps et les ressources nécessaires.

Prévoir une marge d'amélioration

Soyez ouvert aux changements et aux ajustements. Si vous recevez un feedback externe, envisagez de l'intégrer dans le prochain rapport de transparence, le cas échéant. Les feedbacks internes et externes peuvent permettre d'identifier les domaines susceptibles d'être améliorés.

4. Quelles sont les informations et les mesures à inclure dans le rapport de transparence concernant le contenu terroriste en ligne (CTEL) ?

Le règlement CTEL explique à l'article 7.3, les **exigences minimales pour les rapports de transparence**, c'est-à-dire ce qui doit être inclus pour se conformer à cette loi de l'UE. Nous présentons ces exigences dans la check-list ci-dessous :

1) **Informations** sur les mesures prises par le fournisseur de services d'hébergement :

- ☑ identifier les contenus terroristes ;
- ☑ supprimer ou désactiver le contenu terroriste ;
- ☑ empêcher la réapparition et la réimportation de matériel en ligne précédemment bloqué (ceci est particulièrement important lorsque des procédures automatisées sont utilisées).

2) **Chiffres** et, le cas échéant, informations supplémentaires concernant le nombre :

- ☑ de contenus à caractère terroriste (sur la base d'injonctions de retrait ou d'autres mesures) ;
- ☑ d'injonctions de retrait qui n'ont pas été suivies d'effet et les informations supplémentaires sur les raisons de cette situation ;
- ☑ de plaintes traitées par le FSH dans le cadre du mécanisme de plainte, ainsi que des informations complémentaires sur l'issue des plaintes ;
- ☑ de cas dans lesquels le fournisseur de services d'hébergement a rétabli le contenu à la suite de la plainte du fournisseur ou de la fournisseuse de contenu ;
- ☑ de procédures judiciaires engagées par le FSH et les informations complémentaires sur l'issue de ces procédures ;
- ☑ de cas dans lesquels le FSH a dû rétablir le contenu à la suite d'une procédure judiciaire.

C. Merci de votre aide pour lutter contre la menace terroriste !

Félicitations !

Vous êtes arrivé jusqu'ici, ce qui signifie que vous avez acquis les connaissances essentielles sur les exigences du règlement CTEL et sur d'autres mesures visant à lutter contre les contenus terroristes et autres contenus préjudiciables en ligne. Vous contribuerez ainsi à rendre l'internet plus sûr.
Nous sommes conscientes que la mise en œuvre de ces mesures nécessite un investissement et des ressources considérables. Si ce guide vous aide à réfléchir à une stratégie de mise en œuvre des mesures de lutte contre les contenus terroristes adaptée à votre fournisseur de services d'hébergement, alors nous avons déjà beaucoup progressé ! Nous sommes convaincues que ce guide, ainsi que nos autres ressources informatives, est un outil précieux pour permettre à votre fournisseur de services d'hébergement de remplir pleinement ses obligations en matière de lutte contre la menace terroriste en ligne.

Merci beaucoup pour votre engagement dans la lutte contre la menace terroriste en ligne, ainsi que pour la défense de votre fournisseur de services d'hébergement et de vos utilisateurs et utilisatrices.

Note : Le projet TATE comprend également un **cours en ligne gratuit.** Dans ce cours, vous pouvez approfondir le sujet et trouver des détails sur le règlement, des exemples de la façon dont d'autres plateformes mettent en œuvre des mesures spécifiques, et des informations plus générales sur le comportement terroriste en ligne. **Après avoir suivi le cours avec succès, vous recevrez un certificat officiel délivré par des universités renommées** (LMU Munich, Université de Gand (Belgique)). En outre, TATE propose un programme de renforcement des capacités dans le cadre duquel les fournisseurs de services d'hébergement peuvent bénéficier d'un soutien pratique pour répondre aux exigences liées au règlement CTEL.

D. Glossaire

Termes	Explications
Autorités compétentes	Les autorités d'un État membre de l'UE qui sont responsables de la mise en œuvre du règlement CTEL. Une vue d'ensemble des autorités compétentes respectives des États membres est disponible ici.
Fournisseur/fournisseuse de contenu	La personne qui fournit du contenu sur la plateforme concernée, par exemple en publiant un article.
FSH	Fournisseurs de services d'hébergement ; le règlement CTEL s'applique aux fournisseurs de services d'hébergement. Vous trouverez plus de détails sur les fournisseurs de services d'hébergement entrant dans le champ d'application du règlement CTEL ►ici.
PERCI	Le PERCI est un outil coordonné par Europol qui vise à améliorer et à faciliter la communication entre les FSH et les autorités compétentes.
Injonction de retrait	Il s'agit d'une demande qu'un FSH reçoit d'une autorité compétente. Elle informe le fournisseur de services d'hébergement qu'un contenu terroriste a été diffusé sur la plateforme et l'oblige à le retirer rapidement, dans un délai d'une heure à compter de sa réception.
Règlement CTEL (également : LEX 2021/784 & Règlement relatif à la lutte contre la diffusion des contenus à caractère terroriste en ligne)	Le règlement de l'UE contre la diffusion de contenus terroristes en ligne est entré en vigueur en 2022. Il s'applique aux fournisseurs de services d'hébergement (FSH) qui offrent leurs services dans l'UE.
Contenu terroriste en ligne (CTEL)	Contenu comprenant des éléments terroristes ou destiné à promouvoir des objectifs terroristes. Vous trouverez une définition détaillée dans l'►introduction. Le contenu terroriste en ligne est étroitement lié aux ►infractions terroristes, qui peuvent être de nature très diverse.
Conditions générales	Règles contraignantes établies par et pour la plateforme individuelle qui (1) définissent la portée et la responsabilité du FSH envers les utilisateurs et (2) les pratiques d'utilisation appropriées et autorisées, mais aussi interdites. Les utilisateurs et utilisatrices doivent s'y conformer s'ils et elles souhaitent continuer à utiliser le service offert par le FSH. Il existe de nombreux synonymes des conditions générales, comme les conditions d'utilisation et les standards de la communauté.