Organisation de Coopération et de Développement Économiques

DSTI/CDEP(2019)15/FINAL

Non classifié

Français - Or. Anglais

5 août 2021

DIRECTION DE LA SCIENCE, DE LA TECHNOLOGIE ET DE L'INNOVATION
COMITÉ DE LA POLITIQUE DE L'ÉCONOMIE NUMÉRIQUE

Approches actuelles des 50 principaux services mondiaux de partage de contenus en ligne face aux contenus terroristes et extrémistes violents

JT03479913

Ce document, ainsi que les données et cartes qu'il peut comprendre, sont sans préjudice du statut de tout territoire, de la souveraineté s'exerçant sur ce dernier, du tracé des frontières et limites internationales, et du nom de tout territoire, ville ou région.

Table des matières

1 Périmètre, méthodologie et plan de recherche 8

2 Approches des services en matière de lutte contre les contenus terroristes et extrémistes violents : points communs, progrès et tendances 10

Des descriptions vagues des contenus terroristes et extrémistes violents et des concepts connexes, et des approches divergentes pour identifier les « organisations terroristes » 10
Les rapports de transparence portant expressément sur les contenus terroristes et extrémistes violents sont rares 11
Différences entre les rapports de transparence publiés à l'heure actuelle sur les contenus terroristes et extrémistes violents 12
Modérateurs internes, utilisateurs modérateurs et outils automatisés 14
Mécanismes et procédures de notification, de sanction et de recours 15
Divulgation d'informations par les plateformes chinoises 15

3 GIFCT 17

4 Lois et réglementations sur les contenus terroristes et extrémistes violents diffusés en ligne adoptées ou à l'étude 20

Australie 20
Union européenne 22
France 23
Allemagne 24
Corée 25
Royaume-Uni 25
États-Unis 26

Annexe A. Les 50 services de partage de contenus en ligne les plus utilisés au monde 28

Annexe B. Profil des 50 principaux services 34

Annexe C. Définitions 156

Références 157

Résumé

Les terroristes et les extrémistes violents détournent Internet pour servir leurs objectifs. Ils utilisent des applications, des réseaux sociaux et d'autres services de partage de contenus en ligne pour diffuser des messages de propagande faisant l'apologie du terrorisme et de la violence, et favorisant la radicalisation et le recrutement d'individus. Lorsque les contenus terroristes et extrémistes violents peuvent être rapidement reproduits et diffusés en ligne, à moindre coût, les idéologies qu'ils véhiculent se répandent plus facilement. Le présent rapport traite des règles et procédures que les 50 plus grandes plateformes mondiales de partage de contenus en ligne ont mises en place pour lutter contre les contenus terroristes et extrémistes violents, en mettant l'accent sur la question de la transparence.

La propagation des contenus terroristes et extrémistes violents s'est traduite par une multiplication des attaques, dont certaines sont devenues virales, poussant les services de partage de contenus à intensifier leurs efforts pour bannir ces contenus et empêcher leur diffusion. Pour autant, sans un degré de transparence et de responsabilité suffisant, il peut être difficile de comprendre l'approche qu'ils adoptent pour assurer la modération des contenus publiés en ligne et d'évaluer l'efficacité des méthodes employées pour les endiguer. Sans compter qu'ils peuvent involontairement porter atteinte aux droits fondamentaux, notamment aux libertés d'expression et d'accès à l'information, et au droit à un procès équitable.

C'est pourquoi l'Appel de Christchurch (Christchurch Call, 2019[1]), dans lequel les signataires s'engagent à titre volontaire à lutter contre l'utilisation d'Internet à des fins de terrorisme et d'extrémisme violent, reconnaît que les fournisseurs de services en ligne devraient faire preuve de transparence quant aux mesures prises pour prévenir la publication de ces contenus en ligne, ainsi qu'à leurs règles communautaires et conditions générales d'utilisation. On y stipule également que les fournisseurs de services devraient appliquer ces règles et conditions dans le respect des droits humains et des libertés fondamentales, et « effectuer des rapports publics, réguliers et transparents, quantitatifs et reposant sur une méthodologie précise, sur le volume et la nature de contenus terroristes et extrémistes violents détectés et retirés » (Christchurch Call, 2019[1]). La nécessité d'intensifier les efforts pour lutter contre la diffusion des contenus terroristes et extrémistes violents dans des conditions de transparence et de responsabilité et dans le respect des droits et libertés fondamentaux a également été énoncée dans la Déclaration de Hambourg des dirigeants du G20 de 2017 sur la lutte contre le terrorisme (G20, 2017[2]), la Déclaration d'Osaka des dirigeants du G20 de 2019 sur la prévention de l'utilisation d'Internet à des fins de terrorisme et d'extrémisme violent pouvant mener au terrorisme (G20, 2019[3]), et le Résumé de la Présidence de la réunion de 2019 des ministres du numérique du G7 (G7, 2019[4]).

Le présent rapport d'analyse comparative (ci-après dénommé le « rapport ») fait partie de la réponse de l'OCDE à ces appels à l'action. Il dresse la synthèse des pratiques et procédures mises en œuvre par les 50 services de partage de contenus en ligne les plus prisés au monde (ci-après dénommés les « services ») pour lutter contre les contenus terroristes et extrémistes violents ; ce faisant, il en recense les points communs, dégage des tendances quant à leurs approches et indique ceux qui publient des rapports de transparence sur ce type de contenus. Le rapport propose un état des lieux objectif et factuel, et n'a pas vocation à énoncer des recommandations. Il fournit des données probantes à même d'aider à mieux appréhender les règles et procédures mises en place par les services pour lutter contre les contenus terroristes et extrémistes violents et à évaluer le degré de transparence et de responsabilité dont ils font preuve à cet égard. Les conclusions qui s'en dégagent peuvent éclairer utilement les efforts déployés pour apporter une réponse intersectorielle efficace à ce type de contenus, ainsi que les débats y afférents.

Il apporte également de la matière aux activités menées par l'OCDE, en collaboration avec les pays membres, les entreprises, la société civile et les milieux universitaires, en vue de concevoir un cadre multipartite, fondé sur le consensus, et un ensemble d'indicateurs pour les rapports de transparence volontaires que les entreprises publient sur les contenus terroristes et extrémistes violents diffusés en ligne. Le cadre et les indicateurs ont vocation à être intégrés à un modèle normalisé que toutes les

entreprises souhaitant établir des rapports de ce type pourront utiliser, et que tous les Membres de l'OCDE pourront accepter.

Les conclusions phares du rapport sont les suivantes :

- Seuls cinq des 50 principaux services de partage de contenus en ligne publient des rapports de transparence dédiés aux contenus terroristes et extrémistes violents.
- Ils suivent ce faisant des approches d'une grande hétérogénéité. Ils se fondent sur des définitions différentes du terrorisme et de l'extrémisme violent, ne signalent pas les mêmes types d'informations, s'appuient sur des mesures et des méthodes d'estimation disparates et publient leurs rapports à des fréquences et selon des calendriers distincts.
- Compte tenu du nombre peu élevé d'entreprises publiant des rapports et de l'hétérogénéité de la teneur, des calendriers et des modalités y afférents, il est impossible d'avoir une vision intersectorielle claire et exhaustive de l'efficacité des mesures prises par les entreprises pour lutter contre les contenus terroristes et extrémistes violents diffusés en ligne, ou des conséquences de ces mesures sur le respect des droits humains. Cette situation pourrait s'améliorer si les entreprises étaient plus nombreuses à publier des rapports de transparence et y faisaient figurer des informations plus aisément comparables.
- Treize des 50 principaux services de partage de contenus en ligne sont chinois ; aucun d'entre eux ne publie de rapport de transparence sur les contenus terroristes et extrémistes violents diffusés en ligne. Ce qui ne signifie pas qu'il n'y aurait pas d'éléments intéressants à y porter. Les réglementations chinoises interdisent en effet aux fournisseurs et aux éditeurs de contenus en ligne d'afficher des contenus terroristes ou extrémistes. La Chine a adopté en 2017 une loi sur la cybersécurité qui exige que la transmission de contenus interdits soit « immédiatement stoppée » et oblige les entreprises Internet à prêter leur concours aux agences de sécurité dans leurs enquêtes. Le manque de transparence, allié aux obligations réglementaires relatives à la gestion des contenus terroristes et extrémistes violents, peut créer des tensions, à l'heure où les services chinois cherchent à étendre leurs activités dans les pays de l'OCDE.

Introduction

La montée en puissance d'Internet et l'essor des technologies de l'information et des communications ont radicalement transformé la manière dont les individus communiquent, accèdent à l'information et la partagent. Si ce phénomène donne lieu à de nombreux avantages, il s'accompagne également de nouveaux défis et menaces. En effet, Internet et ses technologies sous-jacentes sont de plus en plus utilisés à des fins terroristes (*utilisation d'Internet à des fins terroristes*). Bien qu'il n'existe pas de définition universellement admise du terrorisme ou de l'extrémisme violent, ni, par extension, des contenus terroristes et extrémistes violents, on sait que les terroristes et les extrémistes violents utilisent des applications, des sites de réseaux sociaux et d'autres services de partage de contenus en ligne pour communiquer et coordonner leurs actions dans le monde entier, ainsi que pour diffuser leurs contenus[1], y compris des messages de propagande faisant l'apologie du terrorisme et de la violence et favorisant la radicalisation et le recrutement d'individus[2]. Les contenus terroristes et extrémistes violents publiés en ligne sont d'autant plus préoccupants que l'information numérique peut être reproduite et transmise à un coût quasiment nul. Ils se prêtent par conséquent à une large diffusion et peuvent même devenir « viraux » en très peu de temps, ce qui facilite la propagation des idéologies et de la propagande terroristes et extrémistes violentes. Malheureusement, cette utilisation abusive des services de partage de contenus en ligne a permis aux terroristes et aux extrémistes violents d'entrer en contact, de se développer, de s'organiser, d'agir avec plus de facilité et de rapidité, et d'étendre leur champ d'action.

Face à la prolifération des contenus terroristes et extrémistes violents diffusés en ligne, de grandes entreprises comme Facebook, Twitter et Google, ainsi que certains services de partage de contenus de taille plus modeste, ont dû, sous la pression du public, prendre des mesures pour empêcher les groupes terroristes et extrémistes violents d'exploiter abusivement leurs services. La « modération des contenus » est l'un des principaux outils dont elles disposent. L'expression désigne généralement « la pratique organisée de filtrage des contenus générés par les utilisateurs (CGU) et publiés sur des sites Internet, des réseaux sociaux et d'autres ressources en ligne, afin de déterminer s'ils sont conformes aux règles d'un site, d'une localité ou d'une juridiction donnés »[3] (Roberts, 2017[5]). Le cas échéant, lorsque des contenus sont considérés comme enfreignant les normes de contenu d'une entreprise, ses Règles de communauté ou les cadres juridiques locaux, des mesures peuvent être prises en vue de les retirer ou de les bloquer, de suspendre le compte de l'utilisateur en attendant un examen plus approfondi, ou de procéder à son exclusion définitive de la plateforme. En revanche, sans un degré suffisant de transparence et de reddition de comptes à l'égard de leurs Conditions générales et des utilisateurs, la modération des contenus peut contrevenir aux droits et libertés fondamentaux, tels que la liberté d'expression, l'accès à l'information et le droit à un procès équitable.

Le 15 mars 2019, un homme armé a perpétré et diffusé en direct sur Facebook un attentat terroriste dans deux mosquées de la ville de Christchurch, en Nouvelle-Zélande. Lors de cette attaque, 51 personnes ont été tuées, 50 autres blessées, et la vidéo a été visionnée environ 4 000 fois avant d'être retirée de la plateforme (Christchurch Call, 2019[1]). Elle s'est propagée de manière virale et est restée accessible sur Internet malgré les mesures prises pour la supprimer, mettant au jour la nécessité pour les pouvoirs publics comme pour les entreprises technologiques d'intensifier leurs efforts afin de coordonner et de mettre en œuvre des actions collectives pour éliminer les contenus terroristes et extrémistes violents diffusés en ligne (Christchurch Call, 2019[1]). Deux mois plus tard, le 15 mai 2019, un groupe de chefs de gouvernement et des principaux fournisseurs de services en ligne a lancé l'Appel de Christchurch (2019[1]), dans lequel ils se sont engagés, à titre volontaire, à éradiquer les contenus terroristes et extrémistes violents diffusés en ligne et, ce faisant, à prévenir les utilisations abusives d'Internet par des terroristes et des extrémistes violents.

L'Appel de Christchurch énonce un certain nombre d'engagements forts pris par ses signataires – gouvernements et fournisseurs de services en ligne. Les fournisseurs de services en ligne s'engagent à « [p]rendre des mesures particulières et transparentes permettant de prévenir le téléchargement de

contenus terroristes et extrémistes violents, mais aussi leur diffusion sur les réseaux sociaux et les services analogues de partage de contenus, incluant notamment leur retrait immédiat et permanent », à « [f]aire preuve de plus de transparence dans la mise en place de normes collectives ou de conditions de services », à « [a]ppliquer ces normes collectives ou ces conditions de service dans le respect des droits de l'Homme et des libertés fondamentales » et à « [e]ffectuer des rapports publics, réguliers et transparents, quantitatifs et reposant sur une méthodologie précise, sur la quantité et la nature de contenus terroristes et extrémistes violents détectés et retirés » (Christchurch Call, 2019[1]). Les gouvernements s'engagent quant à eux à agir de concert avec les fournisseurs de services en ligne afin de développer des « solutions techniques pour prévenir le téléchargement [et favoriser] la détection et le retrait immédiat de contenus terroristes et extrémistes violents en ligne », ainsi qu'à définir et mettre en œuvre des « bonnes pratiques en matière de prévention de la diffusion de contenus terroristes et extrémistes violents en ligne », en respectant et en protégeant à tout moment les droits humains, qui peuvent être indûment mis à mal dans le cadre d'activités commerciales (Christchurch Call, 2019[1]). La nécessité d'intensifier les efforts pour lutter contre la diffusion des contenus terroristes et extrémistes violents de manière transparente, responsable et compatible avec les droits et libertés fondamentaux a également été énoncée dans le cadre d'autres forums internationaux, notamment dans la Déclaration de Hambourg des dirigeants du G20 de 2017 sur la lutte contre le terrorisme (G20, 2017[2]), la Déclaration d'Osaka des dirigeants du G20 de 2019 sur la prévention de l'utilisation d'Internet à des fins de terrorisme et d'extrémisme violent pouvant mener au terrorisme, qui a salué l'engagement des plateformes numériques de publier des rapports réguliers et transparents (G20, 2019[3]), ou encore le Résumé de la Présidence de la réunion de 2019 des ministres du numérique du G7 (G7, 2019[4]).

Le présent rapport d'analyse comparative (ci-après dénommé le « rapport ») fait partie de la réponse de l'OCDE aux appels internationaux à l'action formulés dans ces documents. Il dresse une synthèse des pratiques et procédures mises en œuvre par chacun des 50 services de partage de contenus en ligne les plus prisés au monde (ci-après dénommés les « services ») en vue de lutter contre les contenus terroristes et extrémistes violents. Sur la base de ces informations, il recense les points communs, les progrès et les tendances qui se dégagent des approches adoptées par ces services, et précise s'ils publient des rapports de transparence sur ces contenus et, le cas échéant, quelle en est la portée. Il propose un état des lieux objectif, neutre et factuel des approches des services en matière de lutte contre les contenus terroristes et extrémistes violents. Ce faisant, il contribue à l'établissement d'une base de données susceptible d'aider à mieux appréhender les règles et procédures mises en place à cet effet et à évaluer le degré de transparence et de responsabilité dont les services font preuve à l'égard de leurs propres conditions générales et de leurs utilisateurs. Les conclusions qui s'en dégagent peuvent éclairer les efforts déployés pour apporter une réponse intersectorielle à ce type de contenus et les débats y afférents, dans le respect des normes internationales relatives aux droits humains, dont la liberté d'expression, comme énoncé dans le Résumé de la Présidence de la réunion de 2019 des ministres du numérique du G7 (G7, 2019[4]).

Le rapport s'inscrit également dans le cadre d'une réponse plus large de l'OCDE aux appels à l'action précités pour lutter contre les contenus terroristes et extrémistes violents diffusés en ligne et traduit l'engagement de l'Organisation à travailler de concert avec les entreprises et d'autres parties prenantes. Plus particulièrement, ce rapport oriente les efforts menés par l'OCDE, en collaboration avec les pays Membres, les entreprises, la société civile et les milieux universitaires, en vue de concevoir un cadre multipartite, fondé sur le consensus, et un ensemble d'indicateurs pour les rapports de transparence volontaires que les entreprises publient sur les contenus terroristes et extrémistes violents diffusés en ligne. Le cadre et les indicateurs nourriront l'élaboration d'un modèle normalisé que toutes les entreprises souhaitant établir des rapports de ce type pourront utiliser, et que tous les Membres de l'OCDE pourront accepter.

Il convient de souligner que la menace des contenus terroristes et extrémistes violents diffusés en ligne gagne également les plateformes et les applications d'envergure plus modeste, qui ne disposent pas toujours des ressources humaines, technologiques et financières adéquates, ni de l'expertise requise pour

lutter contre l'exploitation de leurs services et de leurs fonctionnalités à des fins terroristes et extrémistes violentes (Tech Against Terrorism, 2019[6]). Il serait donc indiqué de ne pas limiter les prochains rapports d'analyse comparative et autres travaux similaires aux 50 principaux services, mais d'examiner la situation des services de moindre envergure. Cela permettrait de dresser un tableau plus complet du problème et de se tenir au fait des stratégies des groupes terroristes et extrémistes violents.

La section 1 décrit le champ des travaux de recherche exposés dans le rapport et la méthodologie employée. La section 2 dresse un état des lieux des règles et des efforts mis en œuvre par les 50 principaux services mondiaux pour lutter contre les contenus terroristes et extrémistes violents, et en identifie les enjeux, points communs et tendances. La section 3 donne un aperçu du Forum mondial de l'Internet contre le terrorisme (GIFCT), créé à l'initiative d'un groupe d'entreprises technologiques – incluses dans le rapport – pour lutter contre l'exploitation de leurs plateformes à des fins terroristes et extrémistes violentes. La section 4 propose un tour d'horizon, à l'échelle mondiale, des législations et réglementations actuelles concernant les contenus terroristes et extrémistes violents, ainsi que des propositions législatives ou réglementaires en la matière.

1 Périmètre, méthodologie et plan de recherche

Le présent rapport recense les règles, procédures et pratiques mises en œuvre pour lutter contre les contenus terroristes et extrémistes violents par 50 plateformes de médias sociaux, services de communication en ligne, plateformes de partage de fichiers et autres services en ligne dont les activités permettent le chargement, la publication, le partage et/ou le transfert de contenus numériques et/ou facilitent l'échange de contenus vocaux, de vidéos et de messages ou d'autres types de communications en ligne. Lors de la définition du champ de recherche, il a semblé logique d'inclure les services les plus répandus ou « prisés ». Pour ce faire, on a tenté de mesurer leur popularité à partir d'un indicateur commun. Une tâche qui s'est révélée ardue, compte tenu de la diversité des produits et des services proposés, ainsi que des finalités et des publics ciblés[4]. Par exemple, la popularité des plateformes de médias sociaux peut se mesurer d'après le nombre d'utilisateurs actifs mensuels (UAM). En revanche, cet indicateur n'est pas disponible pour les services de partage de fichiers ni pour les encyclopédies en ligne, dont la popularité peut être évaluée à l'aide d'autres critères.

Pour y remédier, les services ont été divisés en trois catégories :

a. les médias sociaux, les services de vidéo en streaming et les services de communications en ligne ;
b. les services cloud de partage de fichiers ; et
c. une catégorie « autres », qui comprend un service de gestion de contenus et une encyclopédie en ligne.

Au sein de chaque catégorie, les services les plus prisés ont été retenus. Pour déterminer les degrés de popularité dans chacune des catégories, les indicateurs suivants ont été utilisés :

- Les plateformes de médias sociaux, les services de vidéo en streaming et les services de communications en ligne ont été choisis en fonction du nombre d'utilisateurs actifs mensuels (UAM). Cet indicateur, couramment utilisé par les analystes du secteur et les investisseurs pour estimer la popularité et la croissance d'un service[5], permet de classer de manière relativement précise, selon leur taille relative, les services dont le succès repose sur l'engagement des utilisateurs.
- Les services cloud de partage de fichiers ont été choisis d'après leurs parts de marché indicatives, une mesure couramment utilisée pour évaluer le poids d'une entreprise dans un secteur d'activité donné.
- La troisième catégorie comprend deux services importants dont la popularité ne peut pas être évaluée selon les mêmes critères que les deux autres, mais qui présentent toutefois un intérêt tel qu'il justifie leur inclusion. Leur poids a été évalué en fonction de données reflétant leur rayonnement et/ou leur utilisation (part de marché indicative et nombre de pages vues mensuellement).

Une liste des 50 services inclus dans cette analyse comparative a été établie (Annexe A).

Les travaux de recherche ont ensuite été menés en trois grandes étapes. Premièrement, on a mis au point un modèle de profil normalisé regroupant tous les domaines d'informations pertinents au regard du champ

de l'étude. On a établi un profil par service, d'après les conditions générales d'utilisation, les lignes directrices et règles de communauté, les blogs, les contrats de service et autres informations officielles (documents constitutifs)[6] librement accessibles. Les services ont été contactés et disposaient d'un délai raisonnable pour formuler des commentaires sur l'exactitude des profils et fournir toute information complémentaire pertinente.

Deuxièmement, les profils ont été mis à jour à la lumière des réponses des services. Les versions finales des profils figurent à l'Annexe B.

Troisièmement, on a recensé les points communs, les progrès et les tendances qui se dégagent des approches des services en matière de lutte contre les contenus terroristes et extrémistes violents. Les conclusions de ces travaux sont exposées à la section 2 du rapport.

Le rapport rassemble des informations sur les règles et les pratiques des services dans plusieurs domaines importants :

a. la définition de termes comme terroriste/terrorisme, et extrémiste violent/extrémisme violent ;
b. la détection et le retrait des contenus terroristes et extrémistes violents, notamment les règles relatives au contrôle du respect des conditions générales d'utilisation du service, aux retraits et aux sanctions, et l'existence éventuelle de procédures de recours ;
c. les conséquences en cas de non-respect, par les utilisateurs, des conditions générales d'utilisation ou des lignes directrices et règles de communauté ;
d. la publication volontaire de rapports de transparence sur les contenus terroristes et extrémistes violents, y compris leur teneur, la méthodologie utilisée et leur fréquence.

Le rapport sera actualisé dans 12 mois ; on répétera alors les trois étapes susmentionnées, de manière à observer l'évolution des règles et procédures des services relatives aux contenus terroristes et extrémistes violents, et la publication des rapports de transparence y afférents.

2 Approches des services en matière de lutte contre les contenus terroristes et extrémistes violents : points communs, progrès et tendances

Des descriptions vagues des contenus terroristes et extrémistes violents et des concepts connexes, et des approches divergentes pour identifier les « organisations terroristes »

Il n'existe pas de définition universellement admise du terrorisme ou de l'extrémisme violent, ni, par extension, des contenus terroristes et extrémistes violents. Il n'est donc guère surprenant que les services n'utilisent pas l'expression « contenus terroristes et extrémistes violents » et ne définissent pas le terrorisme et l'extrémisme violent de manière uniforme. Pour autant, si la majorité des services interdisent explicitement les contenus qui, à des degrés divers, peuvent être considérés comme terroristes et extrémistes violents ou comme dénotant une utilisation d'Internet à des fins terroristes[7], seuls cinq d'entre eux tentent de donner une définition suffisamment détaillée du terrorisme, de l'extrémisme violent et des notions connexes pour éclairer sur le champ de ces termes, en fournissant au besoin des exemples[8].

Les approches des 45 autres services pour décrire les contenus terroristes et/ou les contenus extrémistes violents présentent des degrés de spécificité divers. Parmi eux, 19 services interdisent expressément l'exploitation de leurs technologies pour servir des objectifs terroristes, en utilisant les termes terroriste/terrorisme, extrémistes violents/extrémisme violent et des expressions similaires[9]. Quinze services assimilent les propos haineux et/ou les contenus violents ou choquants à des contenus terroristes et extrémistes violents[10]. Seize services utilisent des descriptions larges et/ou vagues des comportements interdits, lesquelles peuvent être interprétées comme un surensemble englobant les contenus terroristes et extrémistes violents[11].

Les services suivent également des approches différentes pour déterminer ce qu'est une organisation terroriste, avec, à la clé, des répercussions sur ce qui est considéré comme contenu terroriste et extrémiste violent. Par exemple, Facebook indique que ses Règles de la communauté s'appliquent aux activités et groupes terroristes, à l'échelle régionale et mondiale. Avant novembre 2019, ses rapports de transparence évaluaient uniquement les actions entreprises par Facebook face à la propagande terroriste liée à l'EI, Al-Qaida et leurs groupes affiliés[12]. Facebook a alors étendu les indicateurs utilisés dans ses rapports de manière à inclure les efforts déployés pour lutter contre toutes les organisations terroristes (Facebook, 2020[7]). D'autres entreprises, comme Twitter et Automattic (société-mère de WordPress.com) ne précisent pas les groupes sur lesquels ils centrent leurs efforts pour lutter contre les contenus terroristes et extrémistes violents. Certains services se réfèrent aux listes d'organisations terroristes établies par les gouvernements. YouTube, par exemple, précise que sont considérés comme ne respectant pas ses règles relatives à l'extrémisme violent les contenus produits par des organisations terroristes étrangères répertoriées comme telles par des gouvernements[13], sans indiquer à quel(s) gouvernement(s) particulier(s) il est fait référence[14]. Microsoft, pour sa part, déclare considérer comme terroristes les

contenus publiés par ou pour un groupe figurant sur la Liste récapitulative des sanctions du Conseil de sécurité des Nations Unies[15].

Des définitions plus précises, des délimitations plus claires et des sections explicatives sur ce que les services entendent par contenus terroristes et extrémistes violents sont nécessaires afin de déterminer si les services s'attaquent à des contenus identiques, similaires ou différents, notamment en l'absence d'une définition communément admise de cette notion. De même, sans des indications plus claires sur les groupes que les services ciblent dans le cadre de leurs rapports et de leurs approches, il est difficile de comprendre qui est concerné par leurs efforts de modération et comment. Des travaux de recherche ont révélé des différences de taille dans la façon dont les entreprises abordent différents groupes extrémistes violents et montré que toutes les plateformes ne ciblent pas les mêmes groupes de manière identique[16].

Les rapports de transparence[17] portant expressément sur les contenus terroristes et extrémistes violents sont rares

Les entreprises communiquent peu d'informations sur la façon dont elles filtrent et suppriment des contenus en se fondant sur leurs conditions d'utilisation et leurs politiques en général, et sur leurs règles en matière de lutte antiterrorisme et antiviolence en particulier. Sur les 23 services examinés dans ce rapport qui publient des rapports de transparence[18], seuls cinq (Facebook, YouTube, Instagram, Twitter et Automattic) en consacrent spécifiquement aux contenus terroristes et extrémistes violents[19].

Il va de soi que pour comprendre un problème et évaluer les progrès réalisés dans sa résolution, il est essentiel de disposer d'informations fiables et exhaustives. Ces informations sont également nécessaires pour nourrir des délibérations et des débats productifs. Des rapports de transparence ont été établis dans différents domaines[20] pour servir ces objectifs et d'autres qui leur sont connexes. Bien que récente, la modération des contenus en ligne en est un bon exemple. Internet offre aux individus de nombreuses possibilités d'exprimer leurs opinions, leurs points de vue et leurs convictions, et d'accéder à des informations sur des myriades de sujets. Dans le même temps, les plateformes et services en ligne qui servent d'intermédiaires et portent ces contenus et informations peuvent devenir *de facto* des sentinelles chargées – parfois même tenues légalement – de retirer ou de bloquer les contenus publiés par les utilisateurs. L'établissement de rapports de transparence sur le retrait de contenus par les entreprises de l'Internet donne davantage de visibilité à ce rôle, et aide à déterminer si leurs pratiques contribuent à la réalisation des objectifs d'action des entreprises comme des pouvoirs publics (ou du moins ne les compromettent pas), et si elles respectent les droits et libertés fondamentaux des utilisateurs, notamment la liberté d'expression (New America, s.d.[8]).

Google est devenue en 2010 la première entreprise de l'Internet à publier un rapport de transparence, axé sur les demandes de retraits de contenus émanant de l'administration américaine et de gouvernements non américains (Google, 2010[9]). Les retraits de contenus à la demande des gouvernements, lorsqu'ils ont lieu en coulisse, nuisent à la liberté d'expression et à la libre circulation de l'information. Les rapports de ce type avaient donc pour but de dissuader la censure gouvernementale et de préserver les droits et libertés fondamentaux (Google, 2010[9]). Si d'autres entreprises, à l'image de LinkedIn, Microsoft et Twitter, lui ont emboîté le pas dans les années qui ont suivi, les rapports de transparence sont devenus une pratique plus courante en 2013 (New America, s.d.[10]). Des révélations sur la surveillance exercée par le gouvernment ont alors remis en question la gestion des données privées par les entreprises américaines de l'Internet, entraînant une crise de confiance des consommateurs, à laquelle les grands groupes ont répondu en publiant des rapports détaillés sur les demandes de données émanant des gouvernements (New America, s.d.[10]).

Depuis, de nombreuses entreprises technologiques publient des rapports de transparence décrivant le champ et l'ampleur des demandes de données des utilisateurs émanant des pouvoirs publics. Au fil du temps, certaines entreprises ont étendu le périmètre de leurs rapports de transparence pour inclure des

informations sur des sujets tels que les retraits liés à la propriété intellectuelle, les demandes de modération ou de retrait de contenus émanant des pouvoirs publics et des autorités judiciaires, ou l'exploitation des enfants. Pour autant, à l'heure actuelle, seule une poignée de services publie des rapports de transparence traitant spécifiquement de leurs règles et de leurs pratiques en matière de modération des contenus terroristes et/ou extrémistes violents.

Différences entre les rapports de transparence publiés à l'heure actuelle sur les contenus terroristes et extrémistes violents

Les définitions utilisées et les types d'informations figurant dans les cinq rapports de transparence publiés à l'heure actuelle sur les contenus terroristes et extrémistes violents, ainsi que les calendriers et la fréquence de ces rapports, diffèrent systématiquement. Prenons l'exemple des types d'informations considérés. Le service Automattic, spécialisé dans la gestion de contenus (hébergement de blogs, par exemple) indique dans ses rapports de transparence les signalements de contenus terroristes qu'il reçoit des unités de référence Internet (*Internet Referral Units*, IRU) gouvernementales. Il y mentionne le nombre total de signalements reçus, ainsi que le nombre et le pourcentage total de signalements ayant donné lieu à la suspension de sites. On y retrouve également les chiffres mensuels pour chacune de ces catégories[21].

Twitter fait rapport sur le nombre de demandes reçues des pouvoirs publics à l'échelle mondiale concernant le retrait de contenus violant les règles de la plateforme, y compris les règles antiterroristes. Twitter divulgue également le nombre de comptes signalés suite à des soupçons de violations de ses règles ; le nombre de comptes pour lesquels il a pris des mesures selon six catégories d'infractions ; le nombre de comptes suspendus suite à des violations liées à l'apologie du terrorisme, et le pourcentage des infractions détectées par ses outils internes[22].

YouTube communique le nombre de demandes de retrait de contenu émanant de gouvernements selon six catégories ; le nombre de chaînes retirées, triées selon le motif du retrait ; le nombre de vidéos retirées, selon l'origine de la première détection ; le pourcentage de vidéos d'abord identifiées via les méthodes de signalement automatisé, avec et sans vues ; le pourcentage de vidéos retirées, triées selon le motif du retrait (dont le non-respect des règles de YouTube relatives à l'extrémisme violent et aux propos haineux) ; le nombre et le pourcentage de commentaires retirés, selon l'origine de la première détection (signalement automatisé ou humain)[23].

Facebook rend compte de la prévalence des violations liées à la propagande terroriste ; du nombre de contenus pour lesquels des mesures ont été prises ; du pourcentage de contenus en infraction pour lesquels des mesures ont été prises avant que les utilisateurs ne les signalent ; du nombre de recours contre des décisions prises à l'encontre de contenus spécifiques ; et du nombre de contenus rétablis après avoir été retirés[24]. Instagram rend compte des trois premiers indicateurs[25].

Ces rapports sont évidemment utiles. Ils apportent des éclairages indispensables, notamment sur les efforts de modération et de retrait des contenus terroristes et extrémistes violents déployés par les principaux services, sur le nombre de contenus terroristes et extrémistes violents publiés et sur le nombre de contenus bloqués ou retirés par erreur. En revanche, la forte hétérogénéité des rapports rend les comparaisons et les analyses difficiles.

Par exemple, alors que Twitter et Automattic communiquent le nombre de signalements reçus des pouvoirs publics pour des contenus à caractère terroriste publiés sur leurs plateformes, Facebook, Instagram et YouTube ne fournissent pas d'informations comparables. Or la communication systématique de ces signalements pourrait faciliter l'évaluation des efforts gouvernementaux pour faire appliquer la législation et les politiques en matière de lutte contre les contenus terroristes et extrémistes violents en ligne.

Par ailleurs, tandis que YouTube communique le nombre de vidéos retirées selon l'origine de la première détection (signalement automatisé ou humain, par exemple), ce chiffre n'est pas ventilé par catégorie d'infraction aux règles (nudité, propagande terroriste, etc.)[26]. Il est par conséquent impossible de déterminer le pourcentage ou le volume de contenus liés à de la propagande terroriste retirés par YouTube suite à un signalement humain ou automatisé. YouTube communique bel et bien le nombre de signalements humains selon le motif du retrait, dont l'apologie du terrorisme[27], mais ne précise pas quelle proportion de ces contenus a ensuite été retirée. Cela limite l'analyse des approches de YouTube à l'égard de la propagande terroriste.

De même, Facebook et Instagram, bien qu'ayant déployé des efforts considérables pour améliorer leurs rapports de transparence, communiquent le volume de contenus détectés et pour lesquels des mesures ont été prises avant que des utilisateurs ne les signalent[28], mais ne précisent pas si ces contenus ont été repérés par des outils automatisés, des évaluateurs humains ou d'une autre manière. Twitter, pour sa part, indique le pourcentage de comptes suspendus suite à la détection, par ses outils propriétaires, de violations de ses règles antiterroristes[29], une indication utile sur l'importance de ces outils dans ses efforts de lutte contre le terrorisme.

L'approche de Facebook pour l'établissement de rapports sur la modération des contenus terroristes et extrémistes violents est la plus complète. Elle allie des indicateurs clairement définis, des descriptions utiles des méthodologies de calcul et des explications précisant en quoi les indicateurs aident à comprendre la modération des différents types de contenus[30]. Il convient de souligner que la mesure de la prévalence des contenus liés au terrorisme de Facebook est une estimation reposant sur des échantillons de contenus à l'échelle des différentes composantes de la plateforme, comme les Groupes et les Fils d'actualité. Bien que le service ait de bonnes raisons de procéder ainsi, cette mesure étant une estimation, elle peut s'avérer inexacte. De fait, Facebook a été accusé de sous-estimer la prévalence des contenus terroristes publiés sur sa plateforme (Engineering & Technology, 2019[11]). Or, puisque les contenus n'ont un impact que si des utilisateurs les consultent, il est essentiel de disposer de mesures précises de prévalence pour évaluer les résultats des efforts déployés pour lutter contre les contenus terroristes et extrémistes violents. Facebook n'en reste pas moins le seul service du classement à communiquer une mesure de ce type[31].

Parvenir à une parfaite uniformité des rapports de transparence sur les contenus terroristes et extrémistes violents serait un vœu pieux. La teneur d'un rapport de transparence dépend dans une certaine mesure des cibles et des priorités des efforts de modération des services, qui elles-mêmes varient selon les contenus généralement partagés sur leurs plateformes. Cela pourrait expliquer, par exemple, le fait que Twitter s'intéresse au nombre de comptes, et Facebook à celui des éléments de contenus. YouTube, pour sa part, examine à la fois les commentaires et les comptes[32]. De plus, les entreprises incluses dans le classement offrent des services divers et fonctionnent différemment. Selon le modèle économique, l'accès à certaines informations par l'opérateur du service, et par conséquent, le calcul d'indicateurs particulièrement pertinents dans les rapports de transparence d'autres services sur les contenus terroristes et extrémistes violents, peuvent s'avérer difficiles. Par exemple, une application de communication électronique dotée d'un système de chiffrement de bout en bout, comme Telegram, ne peut pas aisément accéder aux contenus que ses utilisateurs partagent. Il lui serait donc difficile de publier une mesure de prévalence comme Facebook. Néamoins, œuvrer en faveur d'une plus grande harmonisation des rapports de transparence sur les contenus terroristes et extrémistes violents, autant que faire se peut, permettrait de mieux identifier et évaluer l'impact des efforts de modération et de retrait des contenus déployés par les services, tout en développant des bonnes pratiques[33].

Modérateurs internes, utilisateurs modérateurs et outils automatisés

Les modérateurs internes employés par les plateformes, les utilisateurs modérateurs et les outils automatisés peuvent tous être mobilisés pour détecter et retirer les contenus répréhensibles, dont les contenus terroristes et extrémistes violents. Chaque approche présente des avantages et des inconvénients. Ainsi, le choix de l'une d'entre elles ou d'une combinaison implique des compromis que les plateformes doivent faire en fonction du type de service offert, de la taille de leur base d'utilisateurs, de leur savoir-faire technologique, de leurs ressources financières, ainsi que d'autres facteurs.

Les modérateurs internes (y compris les sous-traitants) sont des personnes recrutées pour assurer la surveillance et la modération des contenus sur les plateformes et via les services de l'entreprise qui les emploie. Ces modérateurs humains coûtent généralement plus cher que les deux autres solutions et sont moins rapides que les outils automatisés. Par conséquent, les services comptant une large base d'utilisateurs, comme Facebook et YouTube, allient les trois approches, les modérateurs humains ne pouvant à eux seuls assurer une surveillance efficace et efficiente de l'ensemble des contenus partagés sur ces plateformes. En revanche, les modérateurs internes sont à même de prendre des décisions nuancées, un facteur particulièrement important pour la surveillance des contenus tels que ceux liés au terrorisme et à l'extrémisme violent, pour lesquels il n'existe pas de définition claire ni simple.

Les outils automatisés, quant à eux, sont radicalement différents à bien des égards. Plus rapides et moins onéreux que les modérateurs internes (du moins pour ce qui est des coûts marginaux, les coûts fixes pouvant être élevés), ils offrent une solution plus efficiente pour traiter des volumes importants de contenus. En revanche, ils parviennent moins bien à appréhender les contextes plus subtils et reflètent les biais de leurs concepteurs (OFCOM, 2019[12]). L'intervention de modérateurs humains reste donc nécessaire pour percevoir les nuances et corriger les éventuels biais.

La modération par les utilisateurs, qui n'induit quasiment aucun coût, consiste à externaliser la modération des contenus à un corps d'utilisateurs volontaires (Crawford, 2014[13]). Cette approche est parfois critiquée car des utilisateurs, pour décider qui et quoi signaler, font intervenir leurs partis pris et leurs interprétations personnelles des règles de la communauté. Sans compter que certains utilisateurs peuvent abuser de ces possibilités de modération[34].

La plupart des services confient la détection des infractions à leurs conditions générales d'utilisation et à leurs règles à des modérateurs internes[35]. Dix des 50 services analysés recourent à des systèmes faisant appel à des utilisateurs modérateurs[36]. Au moins[37] 21 services utilisent des outils automatisés pour détecter les infractions à leurs conditions générales d'utilisation et à leurs règles[38].

Comme le montre la section 3, les efforts conjoints en matière d'innovation technologique menés sous l'égide du GIFCT ont donné lieu à la création d'une base de données partagée d'« empreintes numériques » uniques dédiée aux contenus terroristes et extrémistes violents que les membres du GIFCT ont retirés de leurs services en ligne. Dans les grandes lignes, l'empreinte des contenus terroristes et extrémistes violents repérés et retirés par l'un des utilisateurs de cette base de données (soit un membre du Hash Sharing Consortium) est partagée avec les autres membres. Cela leur permet d'automatiser l'identification et la modération de ces contenus sur leurs propres plateformes, voire de les bloquer avant qu'ils ne soient publiés. L'adoption d'outils automatisés utilisant la base de données d'empreintes numériques du GIFCT devrait augmenter à l'avenir, à mesure que de nouvelles entreprises rejoignent le GIFCT et son Hash Sharing Consortium. Amazon, Dropbox, LinkedIn, Pinterest et WhatsApp ont récemment adhéré au GIFCT (GIFCT, 2019[14]). Les présents travaux de recherche couvrent tous les membres du GIFCT. Par ailleurs, neuf des 50 services recensés à l'Annexe A sont membres du Hash Sharing Consortium du GIFCT[39].

Mécanismes et procédures de notification, de sanction et de recours

Sur les 50 services étudiés, 21 ont mis en place des mécanismes pour informer les utilisateurs en cas de suspicion d'infraction à leurs conditions générales d'utilisation et autres documents constitutifs[40]. Par ailleurs, 23 services disposent de procédures de recours contre les décisions découlant de la modération des contenus et d'autres mesures prises en vertu de leurs documents constitutifs[41]. Pour ce qui est d'informer les utilisateurs quant aux sanctions prises à leur endroit ou à l'encontre de leur compte, les services adoptent des approches différentes[42].

Les autres services ne disposent pas de procédures de recours ou ne publient pas d'informations à cet égard. Cela pourrait faire craindre que des décisions de modération ne causent des retraits excessifs de contenus sans notification suffisante, en violation des conditions générales d'utilisation des services.

Pour 22 services, il est difficile de déterminer avec exactitude s'ils procèdent à un examen a priori et/ou a posteriori des contenus pour vérifier le respect de leurs conditions générales et de leurs règles[43]. Certains pourraient hésiter à reconnaître qu'ils mènent des activités de surveillance, ou rester vagues pour éviter des critiques s'ils ne procèdent à aucune surveillance des contenus.

Divulgation d'informations par les plateformes chinoises

Les services chinois fournissent en règle générale peu d'informations sur leurs pratiques et procédures en matière de modération des contenus au regard de leurs conditions d'utilisation et de leurs règles[44]. Hormis TikTok, aucun d'entre eux ne publie de rapport de transparence de quelque sorte que ce soit[45].

Cette tendance pourrait s'expliquer par le cadre réglementaire, qui interdit aux fournisseurs de contenus et aux éditeurs en ligne de publier ou d'afficher des contenus qui, entre autres, enfreignent les lois et réglementations de la République populaire de Chine (ci-après dénommée la « Chine »), portent atteinte à la dignité nationale, ou sont à caractère terroriste ou extrémiste (Baidu, Inc., 2017[15]). Il convient de souligner que la notion de contenu « extrémiste » est plus large que celle de contenu extrémiste *violent*. L'environnement réglementaire en Chine crée un système de responsabilité des intermédiaires en vertu duquel les services de partage de contenus en ligne sont juridiquement responsables du contrôle des contenus (Knockel, 2018[16]). De plus, le gouvernement chinois a imposé des exigences de plus en plus strictes, dans une volonté d'accroître son contrôle sur le trafic et les contenus en ligne. Une nouvelle loi de cybersécurité est entrée en vigueur en juin 2017, qui renforce les obligations de censure (par exemple, la transmission des contenus interdits doit être « immédiatement stoppée »), impose la localisation des données, codifie les exigences d'enregistrement avec les noms réels pour les entreprises de l'Internet et les oblige à coopérer avec les organismes de sécurité dans le cadre de leurs enquêtes (Creemers, 2018[17]). Il est attendu des entreprises qu'elles investissent dans du personnel et des technologies de filtrage pour la modération des contenus et qu'elles se conforment aux règles gouvernementales (Knockel, 2018[16]). Le non-respect de ces obligations peut entraîner la révocation des autorisations de fournir des contenus et autres services en ligne, des amendes et/ou la fermeture des services concernés.

Face aux pressions gouvernementales croissantes, les services chinois ont investi davantage à la fois dans les technologies de filtrage et dans les ressources humaines pour les tâches de modération des contenus. Selon Global Times, média chinois contrôlé par l'État, les entreprises technologiques renforcent leurs équipes de censure et mettent au point des outils basés sur l'intelligence artificielle pour passer en revue des « billions de posts, de messages vocaux, de photos et de vidéos chaque jour », afin de vérifier que leurs contenus sont conformes aux lois et aux réglementations (Zhang, 2018[18]). Les rapports des médias indiquent que la plupart des plateformes chinoises sont équipées d'un filtre par mots-clés qui leur permet de censurer automatiquement les informations sensibles avant qu'elles ne soient publiées. Les organes de censure de l'État actualisent en permanence une liste de mots-clés, qu'ils transmettent aux opérateurs de plateformes (Wang, 2019[19]).

Dans ce contexte, les volumes d'informations limités communiqués par les services chinois sur les activités de modération et de surveillance des contenus semblent être en adéquation avec le cadre réglementaire national. Reconnaître publiquement qu'ils surveillent de près les activités des utilisateurs et suppriment tout contenu violant les lois et réglementations applicables pourrait en effet nuire à leur image dans les domaines du respect de la vie privée et de la liberté de parole. Sans compter que cela mettrait en évidence l'absence, ou du moins le flou, des motifs justifiant les retraits de contenus,. Néanmoins, pour respecter ces lois et réglementations, ils sont tenus d'assurer la modération des contenus en étroite coopération avec le gouvernement. WeChat offre un bon exemple de ces intérêts contradictoires. Le service a déclaré publiquement ne pas analyser ou interférer dans le contenu des conversations des utilisateurs ni l'analyser (Corfield, 2018[20]) ; or, des travaux de recherche ont montré que les contenus et messages sont régulièrement censurés sur WeChat (Ruan, 2016[21]).

Cette dichotomie pourrait s'avérer problématique compte tenu des ambitions d'expansion internationale des services chinois, les publics non chinois pouvant se montrer particulièrement méfiants à l'égard des pratiques de surveillance et de censure dictées par le gouvernement chinois (The Washington Post, 2019[22]). L'exemple de TikTok montre que la pénétration des marchés internationaux exige un alignement sur les normes de transparence occidentales en matière de modération des contenus, de politiques et de pratiques. TikTok, dont la base d'utilisateurs va croissant aux États-Unis et dans d'autres pays de l'OCDE[46], a veillé à ce que ses pratiques en matière de modération des contenus ne reposent pas sur des « sensibilités à l'égard de la Chine », appelant à ne pas confondre TikTok avec sa version chinoise, Douyin (Wired, 2019[23]). En octobre 2019, TikTok a annoncé avoir chargé des experts externes d'évaluer certaines de ses règles de modération de contenus (TikTok, 2019[24]) ; le service a publié en décembre 2019 son premier rapport de transparence, qui fait état de 298 demandes légales d'informations utilisateur et de 26 demandes de retraits de contenus émanant du gouvernement au cours du premier semestre de 2019 (TikTok, 2019[25]). En revanche, le rapport de transparence ne couvre pas les activités de Douyin. Il semble que, du fait de la situation exposée plus haut, les versions nationale et internationale de Tiktok soient soumises à un régime différencié, avec en parallèle une approche plus « ouverte » et transparente pour TikTok, qui s'affranchit des exigences réglementaires chinoises, et une approche traditionnelle plus secrète pour Douyin, en adéquation avec le cadre réglementaire chinois. On observe une dualité similaire entre WeChat et sa version nationale Weixin (Ruan, 2016[21]).

3 GIFCT

En réponse notamment à la pression croissante des gouvernements et du public d'endiguer la propagation des contenus terroristes et extrémistes violents diffusés en ligne, Facebook, Microsoft, Twitter et YouTube ont créé en juillet 2017 le Forum mondial de l'Internet contre le terrorisme (Global Internet Forum to Counter Terrorism, ou GIFCT). Amazon, Dropbox, LinkedIn, Pinterest et WhatsApp les ont rejoints plus tard (GIFCT, 2019[14]).

Le GIFCT a pour mission d'empêcher les terroristes et les extrémistes violents d'exploiter à mauvais escient les plateformes numériques[47]. Il sert les objectifs suivants :

- Améliorer la capacité d'un large panel d'entreprises technologiques, indépendamment et collectivement, d'empêcher l'usage de leurs plateformes numériques par des terroristes et des extrémistes violents, et d'y réagir ;
- Permettre un engagement multipartite autour de l'utilisation abusive d'Internet par des terroristes et des extrémistes violents, et encourager les parties prenantes à respecter des engagements clés conformes à la mission du GIFCT ;
- Encourager ceux qui se consacrent au dialogue civique en ligne et soutenir les efforts visant à trouver des alternatives positives aux messages des terroristes et des extrémistes violents ; et
- Faire progresser la compréhension générale des pratiques terroristes et extrémistes violentes et de leur évolution, notamment le point de croisement de leurs activités en ligne et hors ligne (GIFCT, 2017[26]).

Pour ce faire, le GIFCT s'appuie sur quatre stratégies interdépendantes : les efforts conjoints en matière d'innovation technologique, le partage de connaissances, le financement et la recherche, et un protocole de gestion des incidents liés aux contenus. Les efforts de partage d'informations s'inscrivent dans le cadre de la stratégie d'innovation technologique conjointe et sont centrés sur la conception de technologies partagées utilisables par les acteurs du secteur technologique pour prévenir et endiguer la propagation des contenus terroristes en ligne. Ils ont donné lieu à la création d'une base de données partagée à l'échelle du secteur, rassemblant les « empreintes numériques » uniques des images et des vidéos terroristes connues. Les empreintes numériques des images ou des vidéos y sont stockées dans leur forme brute et ne sont pas liées à une plateforme d'origine ou à des données utilisateur. Il s'agit d'une représentation numérique du contenu original et il n'est pas possible de recréer l'image et/ou la vidéo par rétro-ingénierie. Pour savoir à quoi correspond une empreinte, une plateforme doit trouver une correspondance sur son site. Il revient à chaque entreprise de décider de la façon dont elle exploite la base de données, selon ses conditions générales, le fonctionnement de sa plateforme et l'utilisation qu'elle fait de ses capacités techniques et humaines (GIFCT, 2019[27]). Selon le GIFCT, cette collaboration permet une mise en œuvre plus efficiente des règles antiterroristes de ses membres (GIFCT, 2017[28]). La base de données contient à l'heure actuelle plus de 200 000 empreintes, que les entreprises membres peuvent utiliser pour identifier et supprimer le contenu correspondant enfreignant leurs règles respectives ou, dans certains cas, bloquer des contenus terroristes avant même qu'ils ne soient publiés (GIFCT, 2017[28]).

Les entreprises qui utilisent la base de données des empreintes numériques forment le Hash Sharing Consortium. Microsoft, Facebook, Twitter, YouTube, Ask.fm, Cloudinary, Instagram, JustPaste.it, LinkedIn, Verizon Media, Reddit, Snap et Yellow en sont actuellement membres (GIFCT, 2019[27]).

Le GIFCT a publié en juillet 2019 son premier rapport de transparence, dans lequel il précise certaines de ses activités et expose les progrès et les relations à l'échelle intersectorielle (GIFCT, 2019[27]). Le GIFCT y indique qu'aux fins de la base de données des empreintes numériques, et pour dégager des points de convergence, les entreprises fondatrices ont décidé en 2017 de considérer comme terroristes les contenus liés aux organisations figurant sur la Liste récapitulative des sanctions du Conseil de sécurité des Nations Unies. Elles sont également convenues d'une taxinomie de base des contenus publiés en lien avec ces organisations. Celle-ci contient les libellés suivants, que les entreprises appliquent aux contenus lorsqu'elles ajoutent une empreinte numérique à la base de données partagée :

- Menace crédible imminente : publication d'une menace de violence spécifique, imminente et crédible à l'encontre de non-combattants et/ou d'infrastructures civiles.
- Violence explicite à l'encontre de personnes sans défense : meurtre, exécution, viol, torture ou atteintes graves à l'intégrité physique de personnes sans défense (exploitation de prisonniers, non-combattants évidents ciblés).
- Glorification d'actes terroristes : contenus glorifiant, louant, justifiant ou célébrant des attaques ayant été perpétrées.
- Recrutement et formation : contenus visant à recruter des adeptes, les guider ou leur fournir une formation opérationnelle.
- Contenu lié à l'auteur de l'attaque en Nouvelle-Zélande : compte tenu de la viralité et de la diffusion entre les plateformes du manifeste de l'auteur et de la vidéo de l'attentat de Christchurch, et parce que les autorités néo-zélandaises ont déclaré illégal tout élément de ce type, le GIFCT a créé une « banque de crise »[48] pour limiter la propagation de ces contenus.

Le GIFCT fournit également dans son rapport de transparence des informations sur son initiative en matière de partage d'URL. Les entreprises sont uniquement habilitées à retirer le contenu de source primaire des éléments hébergés sur leurs plateformes, ce qui signifie qu'elles peuvent retirer un post, mais que le lien d'origine et le contenu hébergé restent intacts sur une plateforme tierce. En 2018, Twitter a lancé un programme de partage d'URL vers les plateformes liées aux posts Twitter associés à des contenus terroristes. Le GIFCT a étendu ce programme à partir de janvier 2019 afin de permettre à ses membres de partager en toute sécurité des URL suspectes avec les partenaires du secteur auxquels ces URL appartiennent. Le partage en « one-to-one » permet à la plateforme ainsi notifiée de vérifier le lien à la lumière de ses propres conditions générales d'utilisation afin de déterminer si le contenu les enfreint (GIFCT, 2019[27]).

Lors d'une réunion aux Nations Unies, le 24 septembre 2019, les entreprises technologiques et les dirigeants mondiaux ont annoncé un certain nombre de mesures pour mettre en œuvre l'Appel de Christchurch, dont une réforme du GIFCT pour en faire une structure indépendante chargée de piloter la majorité des travaux du secteur des technologies sur la mise en œuvre de l'Appel. Le GIFCT a désormais le statut d'organisation indépendante à but non lucratif de type 501(c)(3) aux États-Unis (GIFCT, 2019[29]). D'autres réformes sont en cours, dont le recrutement d'un Directeur exécutif indépendant chargé de diriger le GIFCT et de coordonner l'ensemble des opérations, de la gestion centrale à la mise en œuvre des programmes, en passant par les activités de financement. Le GIFCT sera administré par un comité opérationnel dirigé par des représentants du secteur, qui travaillera en étroite collaboration avec un large forum multipartite et un comité consultatif indépendant (GIFCT, 2019[30]). Le comité sera dirigé par un représentant non gouvernemental et comprendra des membres de la société civile, des gouvernements et d'entités intergouvernementales[49].

La réforme du GIFCT a également été l'occasion d'une révision de son mandat. Le GIFCT traite désormais des contenus terroristes *et* des contenus extrémistes violents publiés en ligne ; ses missions sont les suivantes :

- l'investissement dans la nouvelle technologie
- la promotion de visions alternatives et d'interventions positives
- le renforcement de l'inclusivité et de la transparence, avec un engagement multipartite à l'échelle des différentes activités, en plaçant la société civile au cœur de la lutte contre les contenus terroristes et extrémistes violents, et
- la mise en place de groupes de travail axés sur six domaines :
  - la transparence
  - la gestion de crise
  - les cadres juridiques
  - les approches techniques
  - les résultats algorithmiques
  - la recherche universitaire et pratique

En outre, le GIFCT prête son concours à la création du Réseau mondial sur l'extrémisme et la technologie (GNET), qui formera un consortium international d'établissements universitaires et d'experts de premier plan, avec des partenariats clés avec des institutions des États-Unis, du Royaume-Uni, d'Australie, d'Allemagne et de Singapour, en vue d'analyser la lutte contre l'utilisation des plateformes numériques à des fins terroristes et extrémistes violentes et de partager les conclusions de cette analyse.

4 Lois et réglementations sur les contenus terroristes et extrémistes violents diffusés en ligne adoptées ou à l'étude

Puisque les groupes terroristes et extrémistes violents utilisent les services en ligne pour diffuser des informations de propagande et de recrutement, les entreprises technologiques sont soumises à des pressions de plus en plus fortes de la part des gouvernements et des institutions du monde entier pour intensifier leurs efforts dans la lutte contre ces activités. Inquiets de constater que les actions menées jusqu'à présent pour combattre les contenus terroristes et extrémistes violents n'ont pas porté leurs fruits, certains États ont commencé à proposer et à adopter des lois et des réglementations, et à mettre en œuvre d'autres d'initiatives pour endiguer la propagation en ligne de ces contenus. Cette section offre un aperçu de ces réponses et récapitule les dispositions juridiques, lois et réglementations qui s'avèrent particulièrement adaptées pour combattre les contenus terroristes et extrémistes violents et l'utilisation d'Internet à des fins terroristes.

Australie

En réponse aux attentats terroristes de Christchurch, le Parlement australien a adopté la *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (loi), entrée en vigueur le 6 avril 2019 (Gouvernement de l'Australie, Federal Register of Legislation, 2019[31]). Elle introduit dans le code pénal de nouvelles infractions relatives aux contenus violents odieux en ligne.

Les contenus violents et odieux désignent tout contenu audio, visuel ou audiovisuel qui enregistre ou diffuse un comportement violent et odieux, réalisé par l'auteur ou les auteurs de ce comportement (ou un complice), qu'une personne raisonnable considèrerait comme choquant dans le contexte. Un comportement violent et odieux est défini comme un meurtre ou une tentative de meurtre, un acte terroriste, des tortures, un viol ou un enlèvement. Il n'est pas nécessaire que la personne soit reconnue coupable d'infraction pour que son comportement constitue un comportement violent et odieux. Le fait que le contenu violent et odieux ait éventuellement été modifié (avec la superposition d'un autre contenu par exemple) ne fait aucune différence au regard de la Loi. Cependant, si le contenu a été modifié de telle sorte qu'il ne remplit plus les critères d'un contenu violent et odieux (à la suite de corrections appropriées), il n'entrera pas dans le cadre de la Loi.

La Loi considère comme une infraction le fait qu'un fournisseur de services Internet, de services de contenu ou de services d'hébergement ne signale pas à l'Australian Federal Police (AFP) « dans un délai raisonnable » un contenu violent et odieux dont il sait qu'il peut être consulté par l'intermédiaire de ses services, lorsque le comportement sous-jacent se déroule ou s'est déroulé en Australie. La Loi ne définit pas le terme « délai raisonnable ». Toutefois, l'exposé des motifs précise qu'il appartiendra en dernier ressort au juge des faits (un jury, par exemple) d'examiner cette question, en fonction de facteurs tels que le volume du contenu (la fréquence à laquelle il a été publié et re-publié, par exemple) et les capacités et

les ressources du fournisseur de services (en l'occurrence, ses capacités techniques de suppression du contenu).

La Loi considère également comme une infraction le fait qu'un fournisseur de services de contenu ou d'hébergement ne supprime pas rapidement de ses services un contenu violent et odieux auquel il est raisonnablement possible d'accéder en Australie (quel que soit le lieu où les services eux-mêmes se situent). Là encore, il incombe au juge des faits de déterminer si le contenu en question a été supprimé « rapidement » ou non, en fonction de facteurs tels que le type et le volume du contenu et les capacités et les ressources du fournisseur de services.

La Loi habilite en outre le eSafety Commissioner à adresser aux fournisseurs de services de contenu ou d'hébergement des avis pour les avertir que leurs services pourraient être utilisés, au moment de l'émission de l'avis, pour consulter des contenus violents et odieux.

Par ailleurs, l'équipe spéciale australienne de lutte contre les contenus terroristes et extrémistes violents en ligne (Australian Taskforce to Combat Terrorist and Extreme Violent Material Online) a été créée en mars 2019 afin d'apporter aux pouvoirs publics des conseils sur les mesures et engagements pratiques, concrets et efficaces à déployer pour lutter contre la mise en ligne et la diffusion de contenus terroristes et extrémistes violents (Department of the Prime Minister and Cabinet, 2019[32]). Afin de remplir sa mission, l'équipe spéciale a publié un rapport le 30 juin 2019 dans lequel elle indique des actions à mener et des recommandations dans cinq domaines : la prévention, la détection et la suppression, la transparence, la dissuasion et le renforcement des capacités. Certaines des actions et recommandations préconisées prévoient les dispositions suivantes :

a. Les plateformes numériques doivent continuer à développer des solutions techniques visant à prévenir la mise en ligne par l'intermédiaire de leurs services de contenus terroristes et extrémistes violents, et à informer les autorités australiennes de l'avancée de ces solutions.

b. Les plateformes numériques doivent travailler avec d'autres membres du GIFCT pour renforcer le consortium de partage d'empreintes numériques et d'adresses URL dans l'optique de s'aligner, dans la mesure du possible, sur les catégories de contenus violents interdites par d'autres plateformes en vertu de leurs propres règles et conditions d'utilisation, telles que la violence explicite ou les contenus violents ou sanglants.

c. Les plateformes numériques doivent disposer de mécanismes de recours clairs et efficaces qui permettent aux utilisateurs de contester des décisions de modération concernant des contenus terroristes et extrémistes violents.

d. Sous l'égide du Comité australo-néo-zélandais contre le terrorisme, les plateformes numériques et les agences publiques australiennes concernées doivent organiser un « événement test » en 2019-2020, autour d'un scénario qui permettra aux parties d'évaluer si les outils du secteur et les processus publics fonctionnent comme prévu, en particulier à mesure qu'ils évoluent avec la technologie et la hausse des investissements dans la modération des contenus.

e. Le gouvernement australien doit continuer à modifier la législation pour établir un cadre permettant de bloquer les contenus terroristes et extrémistes violents en ligne en cas d'événements graves.

f. Les plateformes numériques doivent publier des rapports (tous les semestres au moins) présentant les efforts déployés pour détecter et supprimer les contenus terroristes et extrémistes violents sur leurs services. Destinés à montrer la nature et la portée des actions menées, ces rapports pourront inclure :

- le nombre d'éléments signalés par les utilisateurs comme susceptibles d'enfreindre les règles contre la promotion du terrorisme ou les contenus extrémistes violents ;
- le nombre total d'éléments supprimés par la plateforme ;
- le nombre et le type d'éléments (par exemple, vidéo, chaîne) de contenu terroriste et de contenu extrémiste violent supprimés par la plateforme ;
- des exemples de contenu signalés comme promouvant le terrorisme ou l'extrémisme violent qui enfreignaient et n'enfreignaient pas les règles de la plateforme ;
- le nombre d'éléments de contenu terroriste ou de contenu extrémiste violent signalés ou détectés par les systèmes de la plateforme ;
- le nombre total d'éléments de contenu terroriste et extrémiste violent qui ont été modérés, répartis en fonction de ceux qui ont été signalés par les utilisateurs, les systèmes et d'autres sources, et le volume total des contenus supprimés ;
- le temps moyen passé à examiner les éléments de contenu terroriste et extrémiste violent signalés et à y répondre, ou le nombre de fois qu'un contenu terroriste ou extrémiste violent signalé a été vu par les utilisateurs avant que des mesures ne soient prises ;
- l'application de contrôles appropriés des diffusions en ligne en direct pour diminuer les risques de diffusion en ligne de contenus terroristes et extrémistes violents (Department of the Prime Minister and Cabinet, 2019[32])[1].

Union européenne

Les mesures adoptées par l'Union européenne pour combattre les contenus illicites en ligne ont évolué au fil du temps, passant des initiatives volontaires à des mesures contraignantes qui sont en cours de négociation sous la forme d'une proposition de règlement destiné à prévenir la diffusion des contenus terroristes en ligne.

La Commission européenne (CE) a lancé le Forum de l'UE sur l'Internet en décembre 2015 dans le but de répondre à l'utilisation abusive d'Internet par des groupes terroristes. Le Forum est né d'un partenariat volontaire entre les ministres de l'Intérieur de l'UE, des représentants de l'industrie Internet et d'autres parties prenantes.

Le Code de conduite de l'UE visant à combattre les discours de haine illégaux en ligne a été institué en mai 2016. Facebook, Twitter, YouTube et Microsoft y ont adhéré dès son lancement, suivis un peu plus tard par Snapchat, Instagram, Dailymotion, Google+ et Jeuxvideo. Il vise à garantir que les demandes de suppression de contenu soient traitées rapidement (Commission européenne, 2019[33]).

La CE a présenté en septembre 2017 une Communication intitulée Lutter contre le contenu illicite en ligne (Commission européenne, 2017[34]), qui a apporté des orientations sur la responsabilité des fournisseurs de services en ligne concernant tous les types de contenu illicite en ligne, tels que définis par les législations nationales et européennes.

En mars 2018, la Recommandation de la CE sur les mesures destinées à lutter, de manière efficace, contre les contenus illicites en ligne (Commission européenne, 2018[35]) a retranscrit l'engagement politique de la Communication sous une forme juridique non contraignante. Elle comprend des

propositions de procédures plus rigoureuses pour supprimer les contenus illicites de manière plus efficace et renforcer la protection contre les contenus terroristes en ligne.

La Proposition de règlement sur la prévention de la diffusion de contenus à caractère terroriste en ligne présentée par la CE en septembre 2018 comportait un certain nombre d'exigences, telles que :

a. l'obligation pour les plateformes de supprimer les contenus liés au terrorisme dans un délai d'une heure après avoir reçu un ordre de suppression,
b. l'obligation pour les fournisseurs de services d'hébergement de prendre des mesures proactives pour retirer les contenus terroristes de leurs services, en déployant notamment des outils de détection automatisés, et
c. des propositions de sanctions d'un montant maximum représentant 4 % de leur chiffre d'affaires global pour les plateformes qui ne respectent pas ces obligations.

La proposition impose en outre aux fournisseurs de services d'hébergement l'obligation de publier des rapports de transparence annuels sur les mesures prises contre la diffusion de contenus à caractère terroriste et prévoit qu'ils doivent comprendre au moins les quatre volets suivants :

a. informations sur les mesures prises par le fournisseur de services d'hébergement pour la détection, l'identification et la suppression des contenus à caractère terroriste ;
b. informations sur les mesures prises par le fournisseur de services d'hébergement pour empêcher la remise en ligne de contenus qui ont été supprimés ou dont l'accès a été bloqué parce qu'ils sont considérés comme revêtant un caractère terroriste ;
c. nombre d'éléments de contenu terroriste qui ont été supprimés ou dont l'accès a été bloqué à la suite, respectivement, d'injonctions de suppression, de signalements ou de mesures proactives ;
d. exposé des procédures de plainte et de leur aboutissement.

Adoptée en novembre 2018, la révision de la Directive sur les services de médias audiovisuels comprenait notamment l'obligation pour les plateformes de partage vidéo de permettre le signalement des contenus à caractère terroriste mis en ligne par leurs utilisateurs.

Le Conseil a adopté une approche générale le 6 décembre 2018. Le 17 avril 2019, le Parlement européen a voté en première lecture la Proposition de règlement sur la prévention de la diffusion de contenus à caractère terroriste en ligne, avec de nombreuses modifications à la proposition initiale. Le texte adopté restreint par exemple la définition de contenu à caractère terroriste, rejette diverses dispositions, telles que le recours à des mesures proactives et l'obligation de répondre aux demandes des unités de référence Internet (IRU), et réduit la portée du règlement au contenu public mis en ligne (la définition de « fournisseur de services d'hébergement » ne s'appliquant plus aux services cloud ou d'infrastructures ni aux services de communication électronique) (Parlement européen, 2019[36]). Le Conseil de l'Union européenne, le Parlement européen et la CE mènent actuellement des négociations en trilogue pour parvenir à un accord sur le texte définitif, qui pourrait modifier certains éléments et la portée des dispositions susmentionnées.

France

Après avoir été frappée par de graves attaques terroristes, en particulier l'attentat contre Charlie Hebdo en janvier 2015 et les attentats à Paris en novembre 2015, la France a pris plusieurs mesures pour lutter contre les contenus terroristes et extrémistes violents en ligne. La Loi de 2004 pour la confiance dans l'économie numérique et d'autres lois de lutte contre le terrorisme comprennent notamment :

a. la création d'une liste noire des sites contenant du contenu incitant au terrorisme ou le cautionnant ;

b. l'obligation pour les hébergeurs de ce contenu de le supprimer ; et
c. l'obligation pour les fournisseurs d'accès à Internet de bloquer les sites contenant du contenu illicite (si celui-ci n'a pas été supprimé dans un délai de 24 heures après la demande de suppression).

Pour freiner la propagation des contenus prônant le terrorisme, la France demande aussi leur retrait des résultats des recherches en ligne (Freedom House, 2018[37]).

En mars 2019, le gouvernement français a présenté au Parlement une proposition de loi visant à lutter contre les contenus haineux sur Internet (Gouvernement français, 2018[38]). Après avoir été modifiée plusieurs fois, la proposition comprend finalement des dispositions qui :

a. obligent les opérateurs des plateformes enregistrant un trafic élevé à retirer les contenus haineux manifestement illicites, dont des contenus incitant à la haine ou à la violence, à caractère raciste ou religieux extrémiste, dans un délai de 24 heures après leur notification, sous peine d'encourir une amende de 1.25 million EUR ;
b. obligent les plateformes à retirer les contenus à caractère terroriste et pédophile dans un délai d'une heure après en avoir reçu la demande par les autorités sous peine d'encourir une amende pouvant atteindre 1.25 million EUR ou 4 % du chiffre d'affaires global réalisé par le fournisseur de réseaux sociaux ou d'autres types de contenus en ligne ;
c. obligent les opérateurs à mettre en place sur leurs plateformes un bouton de notification unique pour permettre aux utilisateurs de signaler un abus ;
d. imposent aux plateformes de disposer des ressources humaines et technologiques appropriées pour pouvoir répondre à leurs obligations ;
e. apportent des informations claires sur les mécanismes de recours à la disposition des victimes de contenus haineux sur Internet ;
f. imposent aux plateformes de désigner un représentant légal en France pour collaborer et assurer la liaison avec les autorités ; et
g. confèrent de nouveaux pouvoirs aux autorités administratives ou de réglementation.

Le Sénat a voté la Loi en mars 2020 et l'Assemblée nationale l'a approuvée en mai 2020. Cependant, le cœur de la Loi a été déclaré anticonstitutionnel par le Conseil constitutionnel le 18 juin 2020, car certaines obligations imposées aux opérateurs des plateformes portaient atteinte à la liberté d'expression et de communication (Conseil Constitutionnel, 2020[39]).

Allemagne

En Allemagne, la réglementation des propos haineux en ligne et d'autres formes de discours illicites reflète l'évolution qui s'est opérée dans le discours social sur Internet et la propagation croissante des contenus haineux illicites, en particulier sur les réseaux sociaux. En 2015, le ministère allemand de la Justice et de la Protection du consommateur a créé une équipe spéciale regroupant des représentants des réseaux sociaux et de la société civile (Leisegang, 2017[40]). Les entreprises qui en font partie se sont engagées à améliorer les mécanismes de signalement, les examens et les délais de retrait des contenus illicites. Ces dispositions volontaires ont apporté une certaine amélioration, mais le gouvernement a estimé qu'il convenait de prendre des mesures supplémentaires.

Cette volonté a conduit à l'adoption par le Bundestag de la loi Netzwerkdurchsetzungsgesetz (NetzDG) en juin 2017, qui porte sur la diffusion en ligne des propos haineux et d'autres contenus illicites, dont les contenus terroristes et extrémistes violents (Deutscher Bundestag, 2017[41]). Selon la Loi, un discours haineux et violent sera jugé illicite uniquement si sa circulation et sa diffusion sont soumises à des

poursuites pénales au titre du code pénal allemand. Les plateformes Internet avec plus de 2 millions de visiteurs sont tenues de disposer de systèmes de signalement des publications haineuses et de supprimer le contenu signalé qui est jugé enfreindre l'un des 22 articles y afférent du code pénal allemand. La Loi est entrée partiellement en vigueur en octobre 2017 et a pris pleinement effet le 1er janvier 2018.

Les plateformes Internet doivent supprimer un contenu « manifestement illicite » dans un délai de 24 heures après avoir été notifiées d'une plainte. Si le contenu n'est pas « manifestement illicite », elles disposent de sept jours pour prendre une décision. Si une plateforme reçoit plus de 100 plaintes pour contenu illicite par an, elle doit publier un rapport de transparence en allemand tous les six mois. Des amendes pouvant atteindre 5 millions EUR pour les individus et 50 millions EUR pour les entreprises sont infligées en cas de « négligence répétée ». Les citoyens peuvent signaler les infractions au ministère fédéral de la Justice.

Jusqu'à présent, seul Facebook a été condamné à une amende au titre de la NetzDG, pour avoir fourni des informations incomplètes dans le rapport de transparence du premier semestre de 2018 sur le nombre de plaintes reçues pour contenu illicite. D'après le ministère fédéral de la Justice, le rapport donnait au grand public une image faussée du volume de contenus illicites et des réponses apportées par le réseau social (Office fédéral allemand de la justice, 2019[42]).

Corée

La Corée dispose de plusieurs lois anti-terroristes recouvrant les contenus en ligne. La législation coréenne permet à la direction d'un organisme public concerné de solliciter la coopération de la direction d'une « institution pertinente » pour supprimer, suspendre et surveiller un contenu suspecté de présenter un caractère terroriste ou extrémiste violent.

En juillet 2016, l'Assemblée générale des Nations Unies a adopté une résolution demandant aux États membres des Nations Unies d'élaborer un plan d'action national pour prévenir l'extrémisme violent. Ainsi, la Corée a développé un plan gouvernemental : le « plan d'action national pour la prévention de l'extrémisme violent » a été adopté par le Comité national de lutte contre le terrorisme en janvier 2018 et présenté aux Nations Unies. Il comprend des dispositifs visant à renforcer la coopération public-privé pour créer un environnement Internet sûr et prévenir l'utilisation abusive d'Internet et des technologies de communication par les groupes terroristes.

Le gouvernement coréen participe aussi à l'initiative Tech Against Terrorism lancée par la Direction exécutive du Comité contre le terrorisme (DECT), qui s'appuie sur des contributions volontaires pour lutter contre le terrorisme et alimenter une plateforme de partage des connaissances en la matière. Celle-ci rassemble et partage en ligne des connaissances pour permettre aux grandes entreprises de transmettre aux petites et moyennes entreprises du secteur des technologies de l'information leur savoir-faire en matière de lutte contre l'utilisation abusive d'Internet par des groupes extrémistes violents.

Royaume-Uni

Au Royaume-Uni, le Terrorist Act 2006 définit le contenu terroriste en ligne, et s'attache en particulier à déterminer si le contenu vise à encourager le terrorisme et la diffusion d'informations à caractère terroriste. Le gouvernement a également modifié la législation en adoptant le Counter-Terrorism and Border Security Act 2019, qui permet de condamner jusqu'à 15 ans de prison les personnes qui ont consulté un contenu terroriste en ligne. Cette modification, qui porte sur les contenus consultés ou diffusés en ligne, a durci l'infraction qui existait déjà.

Publié en octobre 2017, le livre vert *Internet Safety Strategy Green Paper* (Bradley MP, 2018[43]) propose une approche stratégique et coordonnée de la sécurité en ligne et présente les mesures possibles pour

remédier à un certain nombre de préjudices en ligne, tels que le harcèlement, la provocation (trollage), la cyberintimidation, le sexting et les abus en ligne. Les autorités ont rendu publique leur réponse au livre vert en mai 2018.

Le 8 avril 2019, le département du Numérique, de la Culture, des Médias et des Sports du gouvernement britannique et le ministère de l'Intérieur ont publié le livre blanc *Online Harms White Paper* (Gouvernement du Royaume-Uni, 2019[44]). Celui-ci vise à répondre à de nombreux types de préjudices en ligne en proposant notamment un nouveau système de responsabilité et de surveillance des entreprises.

La volonté du gouvernement britannique d'instaurer un nouveau devoir de vigilance réglementaire est au cœur de l'élaboration de ce livre blanc. Au titre de ce devoir de vigilance, les entreprises seront davantage tenues responsables des contenus et des comportements préjudiciables hébergés sur leurs plateformes. Elles devront garantir la mise en place de systèmes et de processus efficaces pour réduire le nombre de préjudices en ligne et y répondre. Une autorité de réglementation indépendante sera chargée de vérifier qu'elles appliquent ce devoir de vigilance.

Le livre blanc propose que l'autorité de réglementation dispose des pouvoirs suivants :

a. prendre des sanctions à l'encontre des entreprises qui ne respectent pas le devoir de vigilance ;
b. établir des codes de pratique définissant les étapes que devront suivre les entreprises dans l'exercice de leur devoir de vigilance ;
c. exiger des entreprises la publication de rapports de transparence annuels et leur imposer de publier des informations supplémentaires sur leurs moyens de surveillance et d'exécution ;
d. inciter les entreprises à améliorer leurs mécanismes de signalement et de plaintes pour garantir qu'ils sont efficaces et faciles à utiliser.

Parmi les autres initiatives ou mesures prévues par le livre blanc, figure la possibilité de publier des codes de pratique provisoires permettant d'orienter la lutte contre l'exploitation et les abus sexuels concernant des enfants et les contenus terroristes en ligne pendant la période transitoire entre la consultation sur le livre blanc et la création de l'autorité de réglementation. Un premier rapport publié le 12 février 2020 a résumé les conclusions de la consultation et précisé certaines mesures, en particulier la volonté des pouvoirs publics de nommer l'Ofcom comme autorité de réglementation. Les codes de pratique provisoires relatifs à la lutte contre l'exploitation et les abus sexuels concernant des enfants et les contenus terroristes en ligne seront publiés avec le rapport complet de la consultation.

États-Unis

L'approche américaine des contenus terroristes et extrémistes violents en ligne est guidée principalement par le premier amendement de la Constitution, selon lequel « le Congrès ne fera aucune loi (…) pour limiter la liberté d'expression ». Il protège globalement un large éventail de discours – y compris des discours odieux ou insultants – et interdit généralement toute restriction ou censure préalables des discours de la part du gouvernement. Celui-ci peut néanmoins interdire des discours qui visent à inciter ou à produire de manière imminente un acte illégal et qui sont susceptibles d'inciter ou de produire un tel acte. Par conséquent, plutôt que de criminaliser les propos haineux ou odieux et les discours incitant à la violence ou faisant l'apologie de causes ou de groupes violents, les États-Unis privilégient la poursuite pénale des activités criminelles prônant la violence et la promotion de discours alternatifs crédibles pour fragiliser et contrer les messages terroristes.

Plusieurs articles du Code américain érigent en infraction des comportements associés à des discours qui soutiennent des actes de violence, dont des actes terroristes. Par exemple, le paragraphe 373 du Code

18 U.S.C. considère comme une infraction à la législation américaine le fait de solliciter, ordonner, inciter ou essayer de convaincre de quelque manière que ce soit une autre personne de commettre un délit impliquant la menace, la tentative ou l'utilisation réelle de la force physique contre un tiers ou une propriété.

Par ailleurs, le paragraphe 2339B du code 18 U.S.C., relatif au soutien aux organisations terroristes étrangères, s'applique aux actes commis sous la direction d'organisations terroristes étrangères désignées comme telles, ou en coopération avec elles, que l'auteur des actes sait être des organisations terroristes.

Selon la législation américaine, les fournisseurs de services en ligne ne sont généralement pas tenus responsables des discours tenus par les utilisateurs de leurs services ni des décisions de modération de contenu, sauf dans des circonstances précises, telles que des infractions au code pénal fédéral (voir la Section 230 de la Communications Decency Act). Le cadre de responsabilité des intermédiaires permet aux fournisseurs de services en ligne de modérer la mise en ligne sur leurs plateformes de types de discours qui ne peuvent pas être interdits par les autorités publiques.

Enfin, les fournisseurs de services en ligne n'ont pas le droit de transmettre aux autorités publiques le contenu des communications électroniques sans le consentement de l'auteur, sauf dans certaines circonstances précises (voir la Stored Communications Act).

Annexe A. Les 50 services de partage de contenus en ligne les plus utilisés au monde

Classement	Nom du service (société-mère)	Utilisateurs actifs mensuels, comptes utilisateurs ou visiteurs uniques (millions)	Type de service	Publication de rapports de transparence sur les contenus terroristes et extrémistes violents	Retours/commentaires indiqués sur le profil
1	Facebook (Facebook, Inc.)	2 320 (janvier 2019) (Kemp, 2019[45])	Réseau social et plateforme de streaming	O	O
2	YouTube (Alphabet, Inc.)	1 900 (janvier 2019) (Kemp, 2019[45])	Plateforme de streaming	O	O
3	WhatsApp (Facebook, Inc.)	1 600 (janvier 2019) (Kemp, 2019[45])	Application de messagerie	N	N
4	Facebook Messenger (Facebook, Inc.)	1 300 (janvier 2019) (Kemp, 2019[45])	Application de messagerie	N	N
5	iMessage/FaceTime (Apple, Inc)	1 300 (janvier 2019) (Elmer-Dewitt, 2019[46])	Applications de messagerie et de dialogue en vidéo	N	N
6	Weixin/WeChat (Tencent Holdings Ltd.)	1 098 (janvier 2019) (Kemp, 2019[47])	Réseau social, partage de contenus, messagerie	N	N
7	Instagram (Facebook, Inc.)	1 000 (janvier 2019) (Kemp, 2019[47])	Réseau social	O	O

8	QQ (Tencent Holdings Ltd.)	807 (janvier 2019) (Kemp, 2019[47])	Messagerie instantanée et portail internet	N	N
9	Youku Tudou (Alibaba Group Holding Limited)	580 (août 2019) (Youku Tudou Inc. (NYSE: YOKU), s.d.[48])	Plateforme de streaming (contenu généré par les utilisateurs et contenu syndiqué)	N	N
10	QZone (Tencent Holdings Ltd.)	531 (janvier 2019) (Kemp, 2019[47])	Réseau social	N	N
11	Tik Tok (ByteDance Technology Co.)	500 (janvier 2019) (Kemp, 2019[47])	Application de vidéos courtes	N	O
12	Weibo (Sina Corp.)	2 320 (janvier 2019) (Kemp, 2019[47])	Réseau social	N	N
13	iQIYI (Baidu, Inc.)	454 (décembre 2018) (Baidu, Inc., 2018[49])	Plateforme de diffusion de vidéos (contenu généré par les utilisateurs et contenu syndiqué)	N	N
14	Reddit (Reddit, Inc.)	430 (octobre 2019) (Murphy, 2019[50])	Site agrégateur d'actualités, classement de contenu internet et plateforme de discussion	N	O
15	Twitter (Twitter, Inc.)	326 (janvier 2019) (Kemp, 2019[47])	Réseau social de micromessages	O	O
16	Douban (Information Technology Company, Inc.)	320 (janvier 2019) (Kemp, 2019[47])	Réseau social	N	N
17	LinkedIn (Microsoft, Inc.)	303 (janvier 2019) (Kemp, 2019[47])	Réseau social professionnel	N*	N
18	Baidu Tieba (Baidu, Inc.)	300 (janvier 2019) (Kemp, 2019[47])	Plateforme de communications en ligne	N	N
19	Skype (Microsoft, Inc.)	300 (janvier 2019)	Applications de dialogue en	N*	N

		(Kemp, 2019[47])	vidéo et d'appels vocaux		
20	Quora (Quora, Inc.)	300 (septembre 2018) (Marketing Land, 2 018[51])	Site de questions-réponses	N	N
21	Snapchat (Snap, Inc.)	287 (janvier 2019) (Kemp, 2019[47])	Réseau social	N	O
22	Viber (Rakuten, Inc.)	260 (janvier 2019) (Kemp, 2019[47])	Application de messagerie	N	O
23	Pinterest (Pinterest, Inc.)	250 (janvier 2019) (Kemp, 2019[47])	Réseau social	N	O
24	Vimeo (Vimeo, Inc.)	240 (septembre 2018) (Bicknell, 2018[52])	Application de streaming	N	O
25	IMO (PageBites, Inc.)	211 (avril 2019) (YY Inc. - IR Site, 2019[53])	Applications de dialogue en vidéo et d'appels vocaux	N	N
26	Telegram (Telegram Messenger LLP)	200 (mars 2018) (Pavel, 2018[54])	Application de messagerie	N	N
27	LINE (Line Corporation)	194 (janvier 2019) (Kemp, 2019[47])	Application de messagerie	N	O
28	Ask.fm (IAC [InterActiveCorp])	160 (août 2018) (Kallas, 2019[55])	Réseau social	N	O
29	Twitch (Amazon.com, Inc.)	140 (février 2019) (Iqbal, 2019[56])	Plateforme de streaming	N*	O
30	Xigua (ByteDance Technology Co.)	121 (décembre 2018) (Yang, 2019[57])	Application de diffusion de courtes vidéos	N	N
31	Tumblr (Automattic, Inc.).	115 (août 2018) (Kallas, 2019[55])	Réseau social et microblog	N	N

32	Flickr (SmugMug, Inc.)	112 (août 2018) (Kallas, 2019[55])	Service d'hébergement d'images et de vidéos	N	N
33	Huoshan (ByteDance Technology Co.)	99 (décembre 2018) (Yang, 2019[57])	Application de diffusion de courtes vidéos	N	N
34	VK (Mail.Ru Group)	97 (août 2018) (Kallas, 2019[55])	Réseau social	N	O
35	YY Live/Huya (YY, Inc.)	90 (décembre 2018) (Baidu, Inc., 2018[49])	Plateforme de diffusion en direct	N	N
36	Medium (A Medium Corporation.)	86 (août 2018) (Wickey, 2018[58])	Plateforme de publication en ligne	N	O
37	Haokan (Baidu, Inc.)	75 (décembre 2018) (Yang, 2019[57])	Application de diffusion de courtes vidéos	N	N
38	Odnoklassniki (Mail.Ru Group)	71 (août 2018) (Kallas, 2019[55])	Réseau social	N	N
39	Discord (Discord, Inc.)	56 (mai 2019) (Vincent, 2019[59])	Plateforme de dialogue	N	N
40	Smule (Smule, Inc.)	52 (juillet 2018) (Solsman, 2018[60])	Plateforme de partage de musique et vidéos générées par les utilisateurs	N	N
41	KaoKao Talk (Daum Kakao Corporation)	50 (janvier 2019) (Statista, 2019[61])	Application de messagerie	N	O
42	Deviantart (DeviantArt, Inc.)	52 (2016) (DeviantArt Media Kit, s.d.[62])	Plateforme d'œuvres d'art, de vidéos et de photographies	N	N
43	Meetup (WeWork Companies, Inc.)	35 (août 2018) (Kallas, 2019[55])	Réseau social reposant sur la communauté d'intérêts	N	N
44	4chan (4chan Community Support	22 (août 2019) (4chan,	Plateforme de partage de	N	N

	LLC)	s.d.[63])	contenus		
45	MySpace (Viant Technology/Meredith Corporation)	15 (avril 2016) (Barr, 2016[64])	Réseau social axé sur la musique	N	N

* Le 15 mai 2019, dans le cadre de l'Appel de Christchurch, Amazon, Facebook, Google, Microsoft et Twitter se sont engagés à « publier régulièrement des rapports de transparence sur la détection et la suppression des contenus terroristes ou extrémistes violents » sur leurs plateformes et leurs services et à « garantir que les données s'appuient sur une méthodologie raisonnable et explicable » (https://blogs.microsoft.com/on-the-issues/2019/05/15/the-christchurch-call-and-steps-to-tackle-terrorist-and-violent-extremist-content/). Amazon et Microsoft n'ont toutefois pas encore publié de rapports de transparence sur les contenus terroristes et extrémistes violents.

Les données sur les utilisateurs actifs mensuels ne sont pas disponibles pour certains autres services de partage de contenus en ligne qui ont été utilisés par des terroristes ou des extrémistes violents, mais les chiffres fournis laissent à penser qu'ils devraient figurer dans la liste des 50 principaux services. Le tableau ci-dessous comprend cinq autres services figurant hors classement. Leur importance étant déterminée par d'autres paramètres que le nombre d'utilisateurs mensuels, il n'est pas possible de les comparer aux services répertoriés dans le tableau précédent. Quoi qu'il en soit, pour les besoins du présent rapport, la composition globale du groupe de ces 50 services s'avère plus importante que le classement de chacun des services.

Nom du service (société-mère)	**Part de marché mondial indicative**	**Type de marché/service**	**Rapport de transparence sur les contenus terroristes et extrémistes violents**	**Retours/commentaires indiqués sur le profil**
Google Drive (Alphabet, Inc.)	34.63 % (octobre 2019) (Datanyze, 2019[65])	Partage de fichiers dans le nuage	N	N
Dropbox (Dropbox, Inc.)	24.08 %(octobre 2019) (Datanyze, 2019[65])	Partage de fichiers dans le nuage	N	N
Microsoft OneDrive (Microsoft, Inc.)	10.95 % (octobre 2019) (Datanyze, 2019[65])	Partage de fichiers dans le nuage	N	N

Nom du service (société-mère)	**Part de marché mondial indicative ou pages vues mensuelles**	**Type de marché/service**	**Rapport de transparence sur les contenus terroristes et extrémistes violents**	**Retours/commentaires indiqués sur le profil**
Wordpress.com (Automattic, Inc.)	60 % (avril 2019) (Kinsta, 2011-2019[66])	Système de gestion de contenu	O	N
Wikipedia (Fondation Wikimedia)	18 milliards de pages vues par mois (janvier 2016) (Pew Research Center, 2016[67]) ; 10e site le plus visité au monde (Alexa, 2019[68])	Encyclopédie en ligne	N	N

Annexe B. **Profil des 50 principaux services**

1 Facebook[1]

1 Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Facebook est toutefois l'un des rares services internet à disposer d'une définition bien précise du terrorisme et des termes associés. Dans la section Standards de la communauté, la partie Violence et comportement criminel comprend la rubrique Individus et organismes dangereux (Facebook, s.d.[69]) qui indique que les organisations ou individus qui revendiquent des objectifs violents ou qui sont impliqués dans des activités violentes ne sont pas les bienvenus sur Facebook. Ces organisations et individus sont définis comme celles et ceux qui sont impliqués dans les activités suivantes : ● les activités terroristes, ● la haine organisée, ● les tueries (y compris les tentatives) ou homicides multiples, ● la traite des personnes, ● la violence ou les activités criminelles organisées. Les contenus soutenant ou faisant l'apologie de groupes, dirigeants ou individus impliqués dans ces activités sont supprimés. Les individus (vivants ou décédés) et les groupes suivants ne peuvent pas maintenir une présence (par exemple, en possédant un compte, une page ou un groupe) sur la plateforme : organisations terroristes, terroristes, organisations animées par la haine (ainsi que leurs dirigeants et leurs membres prééminents) et les meurtriers de masse et de plusieurs homicides. Les organisations terroristes et les terroristes comprennent tout acteur non-étatique qui : ● prend part à des actes de violence intentionnels et prémédités, les défend ou les soutient de manière active, ● cause ou tente de causer la mort, des blessures ou des dommages graves à des civils ou à toute autre personne ne prenant pas directement part aux hostilités dans le cadre d'un conflit armé, ou des dommages matériels graves associés à la mort, des blessures graves ou des dommages graves dont les victimes sont des civils, ● dans le but d'assujettir, d'intimider ou d'influencer une population civile, un gouvernement ou une organisation internationale, ● pour atteindre un objectif politique, religieux ou idéologique.

	Une organisation animée par la haine correspond à toute association de trois personnes ou plus organisées sous un nom ou un symbole et dont l'idéologie, les déclarations ou les actions physiques portent atteinte à des individus en fonction de caractéristiques, notamment l'ethnicité ou l'origine ethnique, l'affiliation religieuse, la nationalité, le sexe, l'orientation sexuelle, une maladie grave ou un handicap. Un homicide est considéré comme une tuerie si au moins trois décès sont à déplorer dans le cadre d'un seul incident. Tout individu ayant commis au moins deux meurtres au cours de plusieurs incidents ou à plusieurs endroits est considéré comme un meurtrier multiple. Facebook interdit les symboles qui représentent l'une des organisations ou l'un des individus cités ci-dessus s'ils ne sont pas partagés avec un contexte condamnant le contenu ou en discutant de façon neutre. Les contenus prônant l'une des organisations ou l'un des individus cités ci-dessus ou tout acte commis par eux sont interdits. Facebook interdit la coordination de soutien en faveur des organisations ou individus cités ci-dessus ou de tout acte commis par eux. La plateforme n'autorise pas non plus le contenu qui soutient ou représente des événements qu'elle désigne comme des attaques terroristes, des crimes de haine ou des meurtres de masse. Enfin, dans la rubrique Violence et provocation de la partie Violence et comportement criminel, Facebook (Facebook, s.d.[70]) indique supprimer les messages qui incitent à la violence grave ou la facilitent. Les utilisateurs ne peuvent pas publier : ● Des menaces pouvant entraîner la mort (ou toute autre forme de violence très grave) d'une cible, les menaces correspondant à l'un des éléments suivants : o déclaration de l'intention de commettre un acte de violence très grave, o appel à commettre un acte de violence très grave, y compris les contenus n'indiquant pas de cible précise mais un symbole représentant la cible et/ou l'image d'une arme pour représenter la violence, o déclaration prônant un acte de violence très grave, o déclaration intentionnelle ou conditionnelle de commettre un acte de violence très grave. ● Des contenus qui demandent ou proposent les services de location d'un tueur (par exemple, tueur à gages, mercenaire, assassin) ou préconisent de faire appel à un tueur à gages, un mercenaire ou un assassin contre une cible. ● Des reconnaissances, des déclarations d'intention ou des incitations, des appels à action ou des déclarations intentionnelles ou conditionnelles à enlever une cible. ● Des menaces pouvant causer des blessures graves (violence de gravité moyenne) à des personnes, des personnalités publiques mineures, des personnes ou des groupes vulnérables, les menaces correspondant à l'un des éléments suivants :

	○ déclaration de l'intention de commettre un acte de violence, ○ déclaration faisant l'apologie de la violence, ○ appel à commettre un acte de violence de gravité moyenne, y compris les contenus n'indiquant pas de cible précise mais un symbole représentant la cible, ○ déclaration intentionnelle ou conditionnelle de commettre un acte de violence, ○ contenu concernant d'autres cibles que des personnes, des personnalités publiques mineures, des personnes ou des groupes vulnérables, et tout élément indiqué ci-après considéré comme crédible : ▪ déclaration de l'intention de commettre un acte de violence, ▪ appel à commettre un acte de violence, ▪ déclaration faisant l'apologie de la violence, ▪ déclaration intentionnelle ou conditionnelle de commettre un acte de violence. ● Des menaces pouvant causer un préjudice physique (ou toute autre forme de violence de faible gravité) à des personnes (auto-signalement obligatoire) ou des personnalités publiques mineures, les menaces correspondant à l'un des éléments suivants : ○ déclaration d'intention, ○ appel à action, ○ déclaration intentionnelle ou conditionnelle de commettre un acte de violence de faible gravité ou incitant à le faire. ● Des représentations de personnes ou de personnalités publiques mineures qui ont été manipulées pour contenir des menaces de violence sous forme de texte ou d'image (cible, flèche, pistolet pointé sur la tête, etc.). ● Des contenus créés dans le but exprès de désigner une personne comme un membre appartenant à un groupe à risque précis et reconnaissable. ● Des instructions sur la façon de fabriquer ou d'utiliser des armes s'il existe des preuves de l'intention de blesser gravement ou de tuer des personnes, telles que : ○ des messages énonçant clairement cette intention, ○ des photos ou des vidéos montrant ou simulant le résultat (blessure grave ou mort) dans le cadre des instructions, ○ sauf si les contenus cités ci-dessus sont partagés dans un but d'auto-défense récréative, à des fins d'entraînement militaire, pour des jeux vidéo commerciaux ou pour relater des événements d'actualité (publiés sur une page ou avec le logo « news »).

	● Des instructions sur la façon de fabriquer ou d'utiliser des explosifs, sauf si le contexte indique clairement que le contenu est publié à des fins non violentes (par exemple, jeux vidéo commerciaux, but scientifique/pédagogique clair, feux d'artifice ou matériel de pêche). ● Des contenus comportant des déclarations d'intention, des appels au passage à l'acte ou faisant l'apologie de la violence de forte ou moyenne gravité en raison de la tenue d'élections, de l'inscription d'électeurs ou des résultats d'une élection. ● De fausses informations contribuant à provoquer de la violence imminente ou des préjudices physiques. ● Des appels à action, des déclarations de l'intention d'apporter des armes dans un lieu donné, tel que, mais pas exclusivement, des lieux de culte ou des encouragements à faire de même.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://www.facebook.com/communitystandards/.
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Oui, elles sont disponibles à la page https://about.fb.com/news/2019/05/protecting-live-from-abuse/. Facebook applique en particulier une politique dite de « frappe unique » (one strike) qui interdit les contenus diffusés en direct : quiconque violant les « règles les plus sérieuses » de la plateforme ne pourra plus utiliser la fonction de diffusion en direct (Live) pendant une durée déterminée, 30 jours par exemple, et ce dès la première infraction.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Facebook supprime les contenus de sa plateforme si ces derniers ne respectent pas les Standards de la communauté.
4.1 Notifications des suppressions ou des autres décisions d'exécution	La personne qui a publié le contenu est avertie de la suppression lorsque celle-ci a été effectuée et a la possibilité de demander une révision ou d'accepter la décision (Facebook, s.d.[71]).
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Si l'utilisateur demande une révision, le contenu est soumis à un nouvel examen. Le contenu n'est pas visible sur Facebook tant que celui-ci est en cours. Les évaluateurs ne savent pas que le contenu a déjà fait l'objet d'un premier examen. Les Standards de la communauté n'indiquent pas clairement si le nouvel examen est effectué par une seule personne ou un groupe de personnes, ni quelles formations ou qualifications possèdent les évaluateurs.

	Si l'évaluateur est d'accord avec la décision initiale, le contenu n'est pas republié sur Facebook. En revanche, s'il n'est pas d'accord avec la décision initiale et estime que le contenu n'aurait pas dû être supprimé, celui-ci sera examiné par un troisième évaluateur. C'est la décision de ce dernier qui déterminera si le contenu est autorisé ou non sur Facebook. Pour certains types d'infraction (qui ne sont pas précisés), Facebook autorise la personne qui a publié le contenu à demander un second examen. Lors de ce second examen, le contenu est examiné par des spécialistes de ce type d'infraction et la personne à l'origine du recours a la possibilité de fournir davantage d'informations dans un cadre réservé à cet effet (Facebook, s.d.[71]).
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Facebook s'emploie à trouver un compromis entre transparence et pragmatisme en ce qui concerne les informations communiquées au public, afin de ne pas donner aux terroristes des moyens de contourner ses techniques d'exécution (Facebook, 2018[72]). Il a néanmoins déclaré supprimer les « contenus terroristes et les posts qui soutiennent le terrorisme » dès lors qu'il en a connaissance. (Facebook, 2017[73]) Lorsqu'il reçoit des signalements relatifs à des posts potentiellement terroristes, il les examine avec rigueur et sans délai et, dans les rares cas où il découvre les preuves d'un préjudice imminent, il en informe rapidement les autorités (Facebook, 2017[73]). L'intelligence artificielle figure parmi les outils utilisés par Facebook pour lutter contre le terrorisme, avec notamment des techniques telles que le rapprochement d'images, la compréhension du langage, la suppression de foyers terroristes et la collaboration avec d'autres plateformes (WhatsApp et Instagram en l'occurrence). Facebook recourt depuis quelque temps à l'apprentissage automatique pour détecter des posts qui peuvent exprimer un soutien à l'EIIL ou à Al-Qaida (Facebook, 2018[72]). Facebook note toutefois que l'intelligence artificielle ne permet pas de tout détecter et s'appuie par conséquent également sur l'expertise humaine, dont ses propres utilisateurs (qui peuvent signaler des contenus liés au terrorisme), son équipe « d'opérations de la communauté » (Community Operations team) (Facebook, s.d.[71]), des spécialistes du terrorisme et de la sécurité, ainsi que sur la coopération avec d'autres sociétés telles que Microsoft, Twitter et YouTube, et des institutions publiques et intergouvernementales. La plateforme indique aussi soutenir les programmes de contre-discours avec notamment la Online Civil Courage Initiative (initiative pour le courage civique en ligne) (Facebook, 2017[73])·. Le coût économique marginal d'une utilisation des outils d'intelligence artificielle pour détecter les contenus terroristes et extrémistes violents est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. Facebook est un membre fondateur du GIFCT et participe au consortium de partage d'empreintes numériques.

6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Les conséquences d'une violation des Standards de la communauté varient selon la gravité de la violation et de l'historique de la personne concernée sur la plateforme. Le contenu interdit peut être supprimé. Facebook peut aussi avertir la personne concernée après une première violation, et si la personne continue à enfreindre les règles, la plateforme peut limiter sa capacité à publier des posts ou désactiver son profil. Facebook peut aussi informer les autorités s'il estime qu'un contenu représente un risque réel de préjudice physique ou une atteinte directe à la sécurité publique.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Oui (Facebook, 2018-2019[74]). Facebook publie des rapports de transparence sur l'exécution des Standards de la communauté, qui comportent une partie sur la « propagande terroriste » et une partie sur les « contenus violents et explicites ». Précisons que Facebook indique ne pas tolérer les contenus qui font l'éloge d'organisations terroristes ou de terroristes, les cautionnent ou les représentent. Facebook applique cette règle pour les activités et les groupes terroristes à l'échelle régionale et mondiale. Depuis novembre 2019, la partie consacrée à la propagande terroriste du rapport mesure les actions menées contre toutes les organisations terroristes et non plus uniquement contre la propagande associée à l'EIIL, Al-Qaida et leurs groupes affiliés (Facebook, 2020[7]).
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Publié en novembre 2019, le dernier rapport de transparence comprend les cinq types d'informations suivants dans les parties « propagande terroriste » et « contenus violents et explicites ». - *Prévalence (fréquence des violations relevant de la propagande terroriste et des contenus violents et explicites).* La prévalence désigne le pourcentage de vues des contenus qui enfreignaient certains Standards de la communauté. Facebook estime par exemple que les contenus qui enfreignaient ses règles relatives à la propagande terroriste ont représenté moins de 0.04 % de vues au troisième trimestre 2019. Autrement dit, sur 10 000 contenus vus sur Facebook, moins de quatre comprenaient des éléments considérés par la plateforme comme de la propagande terroriste. (Les chiffres correspondent aux décisions définitives et non aux contenus signalés comme susceptibles de constituer une violation des règles mais qui ont finalement été autorisés.) - *Contenus pour lesquels des mesures ont été prises (quantité de contenus sur lesquels Facebook est intervenu*). Facebook indique qu'un contenu peut désigner « un grand nombre de choses » (Facebook, s.d.[71]), dont un post, une photo, une vidéo ou un commentaire. Les mesures prises peuvent comprendre la suppression d'un contenu, l'ajout d'un filtre d'avertissement sur des photos ou des vidéos susceptibles de déranger certains publics ou la désactivation d'un compte. Ce chiffre indique le nombre total de contenus pour lesquels Facebook a pris des mesures au cours de la période couverte par le rapport parce qu'ils ne respectaient pas les Standards de la communauté. - *Taux de proactivité (taux de contenus pour lesquels Facebook a pris des mesures avant qu'ils n'aient été*

	signalés par des utilisateurs). Ce chiffre indique le pourcentage de contenus pour lesquels Facebook a pris des mesures parce qu'ils enfreignaient ses règles avant qu'ils ne soient signalés par les utilisateurs. Il regroupe les contenus détectés grâce aux outils d'intelligence artificielle utilisés par la plateforme et par les évaluateurs humains. - *Recours (nombre de contenus pour lesquels Facebook a pris des mesures ayant fait l'objet d'un recours).* Ce chiffre indique le nombre total de contenus pour lesquels Facebook a pris des mesures et qui ont fait l'objet d'une demande de ré-examen sur la période couverte par le rapport. - *Contenus rétablis (nombre de contenus que Facebook a remis en ligne après les avoir supprimés).* Ce chiffre indique le nombre total de contenus que Facebook a rétablis après avoir pris des mesures à leur encontre pendant la période couverte par le rapport.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	- *Prévalence.* Ce chiffre indique le nombre estimé de vues d'un contenu enfreignant les règles divisé par le nombre estimé de vues de la totalité des contenus sur Facebook sur la période couverte par le rapport. Par exemple, une prévalence de la propagande terroriste comprise entre 0.18 % et 0.20 % signifie que sur 10 000 contenus vus sur Facebook, 18 à 20 en moyenne ne respectaient pas les règles de la plateforme concernant la propagande terroriste. La prévalence fournit une indication de la fréquence à laquelle est vu un contenu interdit, et non le nombre total de ce type de contenus. Elle est estimée à partir d'échantillons de contenus prélevés dans différentes rubriques de Facebook, telles que les groupes ou le fil d'actualité. Pour les violations relevant de la propagande terroriste, Facebook fournit uniquement une estimation de la limite supérieure, parce qu'il se dit « convaincu que la prévalence des vues des contenus enfreignant ces règles est inférieure à cette limite ». (Facebook, s.d.[71]) Le chiffre de la prévalence inclut les contenus Facebook et Messenger. - *Contenus pour lesquels des mesures ont été prises.* Ce chiffre indique le nombre total de contenus pour lesquels Facebook a pris des mesures au cours de la période couverte par le rapport parce qu'ils ne respectaient pas ses règles. Facebook ne comptabilise pas les contenus qui ont été transmis aux autorités répressives. Ce chiffre regroupe les contenus pour lesquels Facebook a pris des mesures à la suite d'un signalement de la part d'un utilisateur et qui ont été détectés de manière proactive. - *Taux de proactivité.* Ce taux correspond au nombre de contenus pour lesquels Facebook a pris des mesures avant qu'ils ne soient signalés par les utilisateurs divisé par le nombre total de contenus pour lesquels il a pris des mesures. Il inclut les contenus Facebook et Messenger. - *Recours.* Ce chiffre indique le nombre total de contenus pour lesquels Facebook a pris des mesures et qui ont fait l'objet d'une demande de ré-examen sur la période couverte par le rapport. Il inclut les contenus Facebook et Messenger.

	- *Contenus rétablis.* Ce chiffre indique le nombre total de contenus que Facebook a rétablis après avoir pris des mesures à leur encontre pendant la période couverte par le rapport. Facebook peut rétablir un contenu si la décision de le supprimer a fait l'objet d'un recours ou s'il a trouvé une raison de le faire. Il concerne uniquement les contenus Facebook.
10. Fréquence de publication des rapports de transparence.	Facebook indique publier « régulièrement » des rapports de transparence. Il en a jusqu'à présent publié quatre : un pour la période du quatrième trimestre 2017 au premier trimestre 2018 ; un pour la période du deuxième trimestre 2018 au troisième trimestre 2018 ; un pour la période du quatrième trimestre 2018 au premier trimestre 2019, et un pour la période du deuxième trimestre 2019 au troisième trimestre 2019.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Voir les parties 7 et 9 ci-dessus.

2. YouTube

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Le Règlement de la communauté comprend toutefois un certain nombre de précisions applicables aux contenus terroristes et extrémistes violents. Par exemple, les règles concernant les Organisations criminelles et violentes indiquent que les contenus destinés à louer, promouvoir ou aider les organisations criminelles violentes sont interdits sur YouTube. Ces organisations ne sont en outre pas autorisées à utiliser YouTube pour quelque raison que ce soit, y compris le recrutement. Le Règlement ne répertorie pas toutefois les organisations concernées. Il interdit néanmoins les types de contenus suivants : • contenu produit par des organisations criminelles ou terroristes violentes, • contenu faisant l'apologie ou commémorant des personnalités terroristes ou criminelles dans le but d'encourager d'autres personnes à commettre des actes de violence, • contenu louant ou justifiant des actes de violence commis par des organisations criminelles ou terroristes violentes, • contenu destiné à recruter de nouveaux membres pour des organisations criminelles ou terroristes violentes, • contenu présentant des otages, ou mis en ligne dans le but de solliciter, de menacer ou d'intimider des personnes au nom d'une organisation terroriste ou criminelle violente, • contenu exposant les insignes, les logos ou les symboles d'organisations criminelles ou terroristes violentes afin d'en faire l'apologie ou de les promouvoir. Si des contenus liés au terrorisme ou au milieu criminel sont publiés à des fins éducatives, documentaires, scientifiques ou artistiques, ils doivent fournir dans la vidéo ou l'audio suffisamment d'informations pour que les spectateurs

	puissent comprendre le contexte. Les règles concernant les Organisations criminelles et violentes indiquent également les exemples suivants de contenus non autorisés sur YouTube : • remise en ligne brute et non modifiée de contenu créé par des organisations terroristes ou criminelles, • apologie de chefs terroristes ou de leurs crimes à travers des chansons ou des commémorations, • apologie d'organisations terroristes ou criminelles à travers des chansons ou des commémorations, • contenu orientant les utilisateurs vers des sites prônant une idéologie terroriste, diffusant du contenu interdit ou dont le but est de recruter de nouveaux membres. Par ailleurs, les règles concernant les contenus visuels choquants ou violents interdisent les contenus violents ou sanglants destinés à choquer les spectateurs ou à leur inspirer du dégoût ainsi que les contenus incitant à commettre des actes de violence. Elles interdisent notamment les types de contenus suivants : • contenu incitant à commettre des actes de violence contre des individus ou un groupe de personnes en particulier, • vidéos, sons ou images d'accidents de la route, de catastrophes naturelles, de conséquences de guerre/d'attaques terroristes, de combats de rue, d'agressions physiques, d'agressions sexuelles, d'immolation, de torture, de cadavres, de manifestations ou d'émeutes, de vols, d'actes médicaux ou d'autres scènes similaires dont le but est de choquer ou d'inspirer du dégoût aux spectateurs. Les règles concernant l'incitation à la haine interdisent tous les contenus incitant à la violence ou à la haine contre des individus ou des groupes d'individus en fonction de l'une des caractéristiques suivantes : âge, caste, handicap, origine ethnique, identité et expression de genre, nationalité, race, statut d'immigration, religion, sexe/genre, orientation sexuelle, statut de victime d'un événement violent majeur ou de proche d'une victime, statut d'ancien combattant. Les contenus qui encouragent la violence contre des individus ou des groupes d'individus en fonction de l'une des caractéristiques listées ci-dessus ou qui incitent à la haine contre des individus ou des groupes d'individus en fonction de l'une de ces caractéristiques sont interdits. Les exemples fournis pour illustrer les contenus interdits au titre de cette règle citent des phrases qui font l'apologie de la violence contre des individus ou des groupes en fonction de l'une des caractéristiques répertoriées ci-dessus. Enfin, les règles concernant les contenus dangereux ou nuisibles interdisent les contenus qui sont susceptibles d'entraîner des blessures graves ou la mort, c'est-à-dire montrant aux spectateurs comment procéder pour tuer ou mutiler d'autres personnes, tels que des instructions pour fabriquer une bombe destinée à blesser ou à tuer des personnes. Sont également interdits les contenus relatifs à des événements violents faisant la promotion ou l'apologie de tragédies violentes, telles qu'une fusillade dans une école.
2. Manière dont les Conditions de service	Le Règlement de la communauté de YouTube est disponible à l'adresse :https://www.youtube.com/about/policies/#community-guidelines Les règles concernant les organisations criminelles violentes sont disponibles

ou les Règles de la communauté sont communiquées.	à la page :https://www.youtube.com/about/policies/#community-guidelines Les règles concernant les contenus visuels choquants ou violents sont disponibles à la pagehttps://support.google.com/youtube/answer/2802008?hl=fr&ref_topic=9282436 Les règles concernant l'incitation à la haine sont disponibles à la page :https://support.google.com/youtube/answer/2801939?hl=fr&ref_topic=9282436 Les règles concernant les contenus dangereux ou nuisibles sont disponibles à la page :https://support.google.com/youtube/answer/2801964?hl=fr&ref_topic=9282436
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non. Le Règlement de la communauté de YouTube s'applique aux vidéos, aux descriptions des vidéos, aux commentaires, aux diffusions en direct et à tout autre produit ou fonctionnalité YouTube.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Si un contenu enfreint l'une des règles applicables aux contenus YouTube, il est supprimé par YouTube.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Les utilisateurs concernés sont informés de la suppression par courriel ou au moyen d'une notification sur ordinateur ou mobile, ainsi que par une alerte envoyée dans les paramètres de leur chaîne (Google/ YouTube, 2020[75]). Si la suppression d'un contenu fait l'objet d'un avertissement (voir la section 6, ci-après), YouTube indique à l'utilisateur : • le contenu qui a été supprimé, • les règles qui ont été enfreintes, • l'impact de l'avertissement sur sa chaîne, • ce qu'il peut faire ensuite.

4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Si un utilisateur reçoit un avertissement et estime que YouTube a commis une erreur, il peut demander un recours (Google, Youtube, 2020[76]). YouTube informe l'utilisateur des résultats du recours par courriel. Il peut déboucher sur l'un des résultats suivants : • Si YouTube constate que le contenu respecte le Règlement de la communauté, il le réactivera et supprimera l'avertissement de la chaîne de l'utilisateur. Si l'utilisateur a fait appel d'une mise en garde (voir la section 6 ci-après) et que l'appel aboutit, le prochain cas de non-respect du règlement entraînera une mise en garde. • Si YouTube constate que le contenu respecte le Règlement de la communauté, mais ne convient pas à tous les publics, il appliquera une limite d'âge. S'il s'agit d'une vidéo, les utilisateurs non connectés, âgés de moins de 18 ans ou qui ont activé le mode restreint (Google, Youtube, 2020[77]) ne pourront pas la voir. S'il s'agit d'une miniature personnalisée, elle sera supprimée. • Si YouTube constate que la vidéo enfreint effectivement le Règlement de la communauté, il maintiendra l'avertissement sans réactiver la vidéo sur le site. Aucune pénalité supplémentaire n'est appliquée aux appels qui ont été rejetés. Les utilisateurs ne peuvent faire appel d'un avertissement qu'une seule fois.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	YouTube fournit aux utilisateurs des outils permettant de signaler les contenus qui enfreignent le Règlement de la communauté (Google, Youtube, 2020[78]). Il a aussi créé des systèmes automatisés pour détecter les contenus susceptibles d'enfreindre ses règles. Lorsque ces systèmes repèrent un contenu potentiellement problématique, des évaluateurs vérifient s'il enfreint effectivement les règles de la plateforme. Dans l'affirmative, le contenu est supprimé et utilisé pour entraîner les systèmes automatisés et améliorer leurs résultats au fur et à mesure. Concernant plus particulièrement les systèmes automatisés qui détectent les contenus extrémistes (un terme qui n'est pas défini), le personnel de YouTube a examiné manuellement plus de 2 millions de vidéos pour fournir des exemples aux systèmes. YouTube investit par ailleurs dans un réseau rassemblant plus de 180 universitaires, partenaires publics et ONG qui apportent leur expertise aux systèmes d'exécution de la plateforme, en particulier dans le cadre du programme YouTube Trusted Flagger. (Google, Youtube, 2020[79])[2]. Pour l'extrémisme violent, les membres de ce réseau comptent l'International Centre for the Study of Radicalisation du King's College, Londres (The International Centre for the Study of Radicalisation (ICSR), 2020[80]), l'Institute for Strategic Dialogue (ISDGlobal, s.d.[81]), le Wahid Institute en Indonésie et des agences gouvernementales spécialisées dans le contre-terrorisme. Les participants au programme Trusted Flagger reçoivent une formation sur la manière d'appliquer le Règlement de la communauté, et, étant donné que leurs signalements présentent un taux de fiabilité plus élevé que ceux des utilisateurs classiques, ils sont examinés en priorité par les équipes de YouTube. Les contenus signalés par les participants au programme sont soumis aux mêmes règles que les signalements envoyés par d'autres utilisateurs et sont examinés par des équipes à même de décider s'ils enfreignent effectivement le Règlement de la communauté. Les utilisateurs individuels, les agences gouvernementales et les ONG peuvent participer au programme YouTube Trusted Flagger. Les participants doivent s'engager à signaler souvent des contenus susceptibles d'enfreindre le

	Règlement de la communauté et être prêts à échanger régulièrement avec YouTube à propos de divers domaines de contenus de la plateforme. YouTube indique que les discours haineux constituent un domaine où il est difficile d'appliquer des règles à grande échelle, les décisions nécessitant une compréhension nuancée des langues et des contextes locaux. Pour pouvoir mettre en œuvre de manière cohérente ses règles en la matière, YouTube a renforcé les compétences linguistiques et de ce domaine de son équipe d'évaluateurs. Il déploie aussi l'apprentissage automatique pour mieux détecter les contenus potentiellement haineux qui seront ensuite examinés par des humains et en tirer des enseignements susceptibles de s'appliquer à d'autres types de contenus, tels que l'extrémisme violent (Google, Youtube, s.d.[82]). Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus terroristes et extrémistes violents est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. YouTube est un membre fondateur du GIFCT et participe au consortium de partage des empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La première fois qu'un utilisateur publie un contenu qui enfreint les Règles de la communauté, il reçoit une mise en garde et sa chaîne n'est pas sanctionnée. En cas de nouvelle infraction, YouTube envoie un avertissement à l'utilisateur. La chaîne est supprimée si l'utilisateur reçoit trois avertissements en 90 jours. Lors de l'émission du premier avertissement, l'utilisateur ne peut plus effectuer les opérations suivantes pendant une semaine : • mettre en ligne des vidéos, des diffusions en direct ou des stories, • créer des miniatures personnalisées ou des posts destinés à la communauté, • créer ou modifier des playlists, ni y ajouter des collaborateurs, • ajouter ou supprimer des playlists à partir de la page de lecture d'une vidéo via le bouton « Enregistrer ». L'intégralité des droits sont automatiquement restitués après une semaine, mais l'avertissement reste associé à la chaîne de l'utilisateur pendant 90 jours. Si l'utilisateur reçoit un deuxième avertissement dans les 90 jours suivant le premier avertissement, il ne pourra plus publier de contenu pendant deux semaines. S'il ne reçoit aucun nouvel avertissement, ses droits sont automatiquement restitués à l'issue de ces deux semaines, mais chaque avertissement reste associé à sa chaîne pendant 90 jours. La réception de trois avertissements en 90 jours entraîne la suppression définitive de YouTube de la chaîne de l'utilisateur (Google, YouTube, s.d.[83]). Outre ce mécanisme reposant sur trois avertissements, une chaîne peut être clôturée en cas d'abus grave unique (comportement prédateur, par exemple) ou si elle a été créée spécialement pour enfreindre les règles de YouTube (cas des comptes spam, par exemple). Lorsqu'une chaîne est clôturée, toutes ses vidéos sont supprimées. Lorsque des contenus n'enfreignent pas les règles de YouTube, mais s'approchent des critères de suppression et pourraient choquer certains

	spectateurs, certaines fonctionnalités peuvent être désactivées. Il peut s'agir de contenus suivants : • contenus religieux provocateurs ou suprémacistes sans appel direct à la violence ni visant principalement à encourager la haine, • théories conspirationnistes attribuant des intentions maléfiques, malveillantes ou de corruption à des individus ou des groupes en fonction de certaines caractéristiques, • vidéos niant l'existence d'un événement violent bien documenté. Le contenu reste disponible sur YouTube, mais la page de lecture n'affiche plus de commentaires, de suggestions de vidéos ou de mentions « J'aime », et un message d'avertissement s'affiche avant la vidéo. Ces vidéos ne peuvent en outre pas être monétisées. Le fait que certaines fonctionnalités soient désactivées n'entraîne pas l'envoi d'un avertissement à la chaîne (Google, YouTube, s.d.[84]). YouTube informe l'utilisateur de sa décision de désactiver certaines fonctionnalités par courriel. L'utilisateur peut faire appel de cette décision.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Oui (Google, s.d.[85]). YouTube publie des rapports de transparence sur l'exécution du Règlement de la communauté. L'une des parties de ces rapports est consacrée au « Contenu extrémiste violent » (Google, YouTube, s.d.[86]). Le dernier rapport de transparence publié indique que les contenus qui ne respectent pas les règles de YouTube relatives à l'extrémisme violent incluent les contenus produits par des organisations terroristes étrangères répertoriées comme telles par les autorités (sans toutefois préciser de quelles autorités il s'agit). Il indique aussi que YouTube interdit les contenus promouvant le terrorisme, tels que des contenus faisant l'apologie d'actes terroristes ou incitant à la violence. Il mentionne en outre que les contenus produits par des groupes extrémistes violents qui ne sont pas répertoriés comme organisations terroristes étrangères par des autorités sont souvent couverts par les règles de YouTube sur la publication de contenus haineux, violents ou choquants (voir la section 1, ci-dessus), y compris les contenus principalement destinés à choquer, ou à caractère sensationnel ou gratuit.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	YouTube publie : • le nombre de demandes de suppression de contenus émises par les autorités, réparties en six catégories (sécurité nationale, diffamation, biens et services réglementés, confidentialité et sécurité, droits d'auteur, tous les autres) (Google, 2010-2019[87]) ; • le nombre de chaînes clôturées, en fonction du motif de clôture (dont l'incitation à la violence et l'extrémisme violent) ; • le nombre de vidéos supprimées par source de détection initiale (détection automatique, Trusted Flagger individuels, utilisateurs, ONG et autorités administratives) ; • le pourcentage de vidéos détectées initialement par les systèmes automatiques, avec ou sans aucune vue, c'est-à-dire le pourcentage de vidéos qui ont été supprimées avant d'avoir été vues par rapport à celles qui ont été supprimées après avoir obtenu quelques vues ; • le nombre et le pourcentage de signalements humains, par motif de signalement (dont l'incitation au terrorisme). YouTube indique qu'une vidéo peut être signalée plusieurs fois et pour différentes raisons et que le signalement n'entraîne pas forcément la suppression. Les vidéos signalées par des personnes sont supprimées lorsqu'un

	modérateur qualifié confirme qu'elles ne respectent pas les règles (Google, 2010-2019[87]) ; • le pourcentage et le nombre de vidéos supprimées selon le motif de suppression (dont le non-respect des règles de YouTube relatives à l'extrémisme violent et des règles concernant l'incitation à la haine) (Google, YouTube, s.d.[86]) ; • le nombre et le pourcentage de commentaires supprimés selon le motif de suppression (dont le non-respect des règles de YouTube relatives à l'extrémisme violent et des règles concernant l'incitation à la haine) ; • le pourcentage de commentaires supprimés en fonction de l'origine du premier signalement (détection automatique ou manuelle) (Google, s.d.[85]).
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Pas d'informations communiquées.
10. Fréquence de publication des rapports de transparence.	Trimestrielle (Google, s.d.[88]).
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Voir les sections 7 et 8 ci-dessus.

3. WhatsApp

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Les Conditions d'utilisation de WhatsApp ne définissent pas les contenus terroristes et extrémistes violents. Cependant, la partie « Sécurité et intégrité » des Conditions d'utilisation indique que WhatsApp s'efforce de garantir la sécurité et l'intégrité de ses services en traitant de manière appropriée les personnes faisant preuve d'un comportement abusif et menant des activités contraires à ses Conditions d'utilisation. Les termes « comportement abusif » et « activités contraires à ses Conditions d'utilisation » peuvent s'appliquer aux utilisateurs diffusant des contenus terroristes et extrémistes violents, bien que ce ne soit pas précisé explicitement. Ces termes ne sont pas définis. Les Conditions d'utilisation interdisent toute utilisation des services à mauvais escient, « tout comportement nuisible envers autrui » et toute violation des Conditions et politiques.

	WhatsApp indique que les utilisateurs doivent consulter et utiliser ses services uniquement à des « fins légales, autorisées et acceptables », notamment ne pas les utiliser d'une manière qui soit de « nature illégale, obscène, diffamatoire, menaçante, intimidante, haineuse, racialement ou ethniquement offensante, assimilée à du harcèlement ou incite ou encourage un comportement illégal ou déplacé pour d'autres raisons, y compris la promotion de crimes violents ».
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://www.whatsapp.com/legal/#terms-of-service
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	WhatsApp indique de façon générale qu'il peut modifier, suspendre ou résilier l'utilisation ou l'accès à ses services à tout moment en cas de comportement suspect ou illégal ou s'il estime raisonnablement que l'utilisateur ne respecte pas ses Conditions d'utilisation ou crée un préjudice ou un risque de préjudice pour les utilisateurs ou d'autres personnes. Aucune procédure de recours n'est indiquée, mais si un utilisateur pense que son compte a été résilié ou suspendu par erreur, il peut contacter WhatsApp par courriel, à l'adresse support@whatsapp.com.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée,
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée, mais si un utilisateur pense que son compte a été résilié ou suspendu par erreur, il peut contacter WhatsApp par courriel, à l'adresse support@whatsapp.com.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	WhatsApp indique qu'il développe des systèmes automatisés pour améliorer sa capacité à détecter et supprimer les comportements abusifs et les activités susceptibles de porter préjudice à sa communauté et à la sécurité et à l'intégrité de ses services. Les utilisateurs peuvent aussi signaler les contenus qu'ils estiment problématiques. Les signalements sont ensuite examinés par les modérateurs de WhatsApp qui prennent des mesures appropriées si nécessaire. Étant donné que WhatsApp fait partie des entités Facebook, il est possible qu'il utilise les mêmes méthodes que ce dernier pour détecter et supprimer les contenus terroristes et violents,

	notamment parce que WhatsApp indique partager des informations avec Facebook pour lutter contre le spam, les menaces, les infractions et promouvoir la sécurité et l'intégrité sur l'ensemble des produits des entités Facebook. Cependant, dans la mesure où les communications sur WhatsApp sont cryptées, il est difficile d'imaginer comment les contenus terroristes et extrémistes violents peuvent être interceptés. Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus terroristes et extrémistes violents est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. WhatsApp est membre du GIFCT depuis peu.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Si un utilisateur ne respecte pas les Conditions d'utilisation et politiques de WhatsApp, celui-ci peut prendre des mesures à l'encontre du compte de l'utilisateur, telles que le désactiver ou le suspendre. Dans ce cas, l'utilisateur ne doit pas créer un autre compte sans y être autorisé par WhatsApp. WhatsApp indique également que s'il a connaissance d'un comportement abusif ou d'activités contraires à ses Conditions d'utilisation, il prendra des mesures appropriées en supprimant le compte ou les activités de l'utilisateur concerné ou en contactant les autorités répressives.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Pas encore, mais la publication des rapports de transparence étant une condition de la participation au GIFCT, l'on peut s'attendre à ce que WhatsApp le fasse dans un futur proche.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Par exemple, après les attentats de Christchurch, deux extrémistes violents d'extrême droite qui faisaient apparemment partie d'un groupe WhatsApp intitulé « Christian White Militia » ont publié des déclarations encourageant le terrorisme en mars 2019 (Dearden, 2019[89]).

4. Facebook Messenger (Facebook, Inc.)

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Facebook Messenger ne possède ni Conditions générales d'utilisation ni Règles de la communauté particulières. Cependant, dans la mesure où Facebook balaie les conversations publiées sur Facebook Messenger pour détecter d'éventuelles violations à ses propres Standards de la communauté (Frier, 2018[90]), ces standards, qui comprennent une description précise du terrorisme et des notions associées, s'appliquent à Facebook Messenger. Voir la section 1 du profil de Facebook.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://www.facebook.com/communitystandards/.
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Voir la section 4 du profil de Facebook.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Voir la section 4.1 du profil de Facebook.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Voir la section 4.2 du profil de Facebook.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs	Voir la section 5 du profil de Facebook.

humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Voir la section 6 du profil de Facebook.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Voir la section 7 du profil de Facebook.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Voir la section 8 du profil de Facebook.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Voir la section 9 du profil de Facebook.
10. Fréquence de publication des rapports de transparence.	Voir la section 10 du profil de Facebook.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Voir les sections 7 et 8 du profil de Facebook.

5. iMessage/FaceTime

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Cependant, les Termes et conditions générales d'utilisation des services médias d'Apple (qui s'appliquent à iMessage et FaceTime) interdisent aux utilisateurs de publier des contenus répréhensibles, insultants, illégaux, trompeurs ou nuisibles, tels que des commentaires, des photos, des vidéos et des podcasts (ainsi que des œuvres d'art et des métadonnées associées).

2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://www.apple.com/ca/legal/internet-services/itunes/ca/terms.html
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Aucune procédure n'est indiquée. Apple indique de manière générale qu'il peut vérifier et décider de supprimer ou de modifier les contenus publiés.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Apple dispose d'un dispositif de signalement permettant aux utilisateurs de signaler les contenus qui enfreignent ses règles de publication (Submission Guidelines, figurant dans ses Termes et conditions générales d'utilisation des services médias). Les signalements sont ensuite vérifiés et traités par l'équipe d'Apple. Dans la mesure où iMessage et FaceTime sont cryptés, il est difficile de comprendre comment un algorithme ou un évaluateur travaillant pour Apple pourrait détecter un contenu problématique et notamment un contenu terroriste et extrémiste violent. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus problématiques est probablement relativement élevé. Apple n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions	Si Apple détecte une violation potentielle ou avérée de l'une des dispositions de ses Conditions générales d'utilisation, il peut sans en avertir l'utilisateur au préalable désactiver son identifiant Apple, sa licence d'utilisation des logiciels Appel et/ou son accès à ses

générales d'utilisation ou des Règles de la communauté.	services, qui comprennent iMessage et FaceTime.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. Apple publie des rapports de transparence (Apple, s.d.[91]) qui comportent une section sur les demandes de suppression de contenu émanant des autorités et de tiers privés signalant des violations de ses Conditions générales d'utilisation ou de la législation locale, mais qui ne mentionnent pas précisément les contenus terroristes et extrémistes violents.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	C'est possible. Un manuel sur la sécurité publié par l'EIIL recommandait d'utiliser iMessage pour protéger l'identité des sympathisants du mouvement (Zetter, 2015[92]), mais rien ne prouve que ces derniers l'ont effectivement utilisé (Dilger, 2015[93]).

6. WeChat

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation interdisent toutefois aux utilisateurs de soumettre, de mettre en ligne ou d'afficher des contenus qui, selon l'avis raisonnable de WeChat ou effectivement : • enfreignent des lois ou des réglementations (ou peuvent entraîner une violation des lois ou des réglementations) ; • créent un risque de perte ou de préjudice pour des personnes ; • nuisent à des personnes (adultes ou mineures) ou les exploitent de quelque façon que ce soit, notamment par des messages d'intimidation, de harcèlement ou des menaces de violence ; • sont haineux, assimilés à du harcèlement, insultants, racialement ou ethniquement offensants, diffamatoires, humiliants (publiquement ou de quelque façon que ce soit), menaçants, grossiers ou répréhensibles de quelque façon que ce soit.
2. Manière dont les Conditions générales d'utilisation ou les	Disponibles aux pages https://www.wechat.com/en/service_terms.html et

Règles de la communauté sont communiquées.	https://www.wechat.com/en/acceptable_use_policy.html (Tencent, s.d.[94])
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	WeChat indique qu'il peut examiner (mais ne s'engage pas à le faire) les contenus (dont les contenus postés par ses utilisateurs) ou les programmes ou services tiers proposés sur WeChat pour vérifier s'ils respectent ses politiques, les lois et réglementations en vigueur ou ne sont pas répréhensibles de quelque manière que ce soit, et se réserve le droit de bloquer ou supprimer des contenus pour quelque raison que ce soit, conformément aux lois et réglementations en vigueur.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Pas d'informations communiquées. Les entreprises chinoises en ligne, dont WeChat, possèderaient une équipe de modérateurs chargée de traiter les contenus problématiques[3]. Des recherches ont également montré que WeChat utilise des algorithmes (Knockel, 2018[16]), ainsi que des dispositifs de filtrage par mots clés et de blocage des adresses URL (Ruan, 2016[21]) pour interdire les contenus qui enfreignent ses Conditions générales d'utilisation (et qui peuvent comprendre des contenus terroristes et extrémistes violents), même s'il appliquerait ces techniques uniquement aux comptes rattachés à des numéros de téléphone enregistrés en Chine continentale (Ruan, 2016[21]). Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus problématiques est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. WeChat n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.

6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	WeChat indique qu'il peut suspendre ou résilier l'accès à ses services s'il pense raisonnablement qu'un utilisateur a enfreint les Conditions générales d'utilisation, que son utilisation crée un risque pour WeChat ou d'autres utilisateurs, que la suspension ou la résiliation est requise par la législation en vigueur ou à sa seule et absolue discrétion.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. La tuerie de Christchurch a été postée sur WeChat (Kenny, 2019[95]). WeChat a été utilisé également pour diffuser de la propagande anti-musulmans (Huang, 2018[96]).

7. Instagram

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Globalement, Facebook et Instagram partagent les mêmes politiques. Facebook indique qu'un contenu considéré comme ne respectant pas ses propres politiques sera également considéré comme tel sur Instagram. Celui-ci applique donc les définitions présentées dans le profil de Facebook (voir la section 1 du profil de Facebook). Les Standards de la communauté de Facebook étant plus complets que les Règles de la communauté d'Instagram, ils servent de référence même lorsqu'il s'agit de traiter des infractions commises sur Instagram. Les Règles de la communauté d'Instagram précisent qu'Instagram n'est pas un espace où soutenir ou faire l'éloge du terrorisme, du crime organisé ou de groupes haineux, ni où encourager la violence ou attaquer quiconque en raison de sa couleur de peau, de son origine ethnique, de sa nationalité, de son sexe, de son genre, de son identité de genre, de son orientation sexuelle, de son affiliation religieuse, de handicaps ou d'états pathologiques. Les menaces sérieuses d'atteintes à la sécurité publique et personnelle sont interdites, de même que le partage d'images visant à glorifier la violence.

2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Les Règles de la communauté d'Instagram sont disponibles à l'adresse :https://help.instagram.com/477434105621119?helpref=page_content Les Conditions d'utilisation d'Instagram sont disponibles à l'adresse :https://help.instagram.com/581066165581870
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Instagram peut supprimer un contenu s'il enfreint les Règles de la communauté, ou désactiver ou résilier un compte.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Instagram avertit l'utilisateur concerné de la suppression d'un contenu ou de la suspension ou de la résiliation de son compte.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Si l'utilisateur pense que son contenu a été supprimé, ou que son compte a été résilié par erreur, il peut faire appel de cette décision. Il est possible de faire appel de la décision de supprimer un contenu considéré comme contraire aux politiques « de lutte contre le terrorisme » d'Instagram (qui ne sont pas précisées). Si le contenu a effectivement été supprimé par erreur, Instagram le remettra en ligne et retirera l'infraction des archives du compte. En février 2020, Instagram a mis en place un mécanisme de recours simplifié pour les comptes désactivés, accessible directement depuis l'application, sans passer par le Centre d'aide. Voir https://about.instagram.com/blog/announcements/safer-internet-day-2020/

5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Instagram dispose d'une option de signalement intégrée qui permet aux utilisateurs de signaler les contenus qui enfreignent les Règles de la communauté. Il possède aussi une équipe mondiale qui examine les signalements et supprime les contenus qui transgressent ses règles. Instagram déclare qu'il peut collaborer avec les autorités, en particulier s'il estime qu'un contenu représente un réel risque de préjudice physique ou une atteinte directe à la sécurité publique. En outre, étant donné qu'il fait partie des « entités Facebook », il peut utiliser les mêmes méthodes pour détecter et supprimer des contenus terroristes et extrémistes violents. Après la tuerie de Christchurch, Sheryl Sandberg, la directrice des opérations de Facebook a publié sur la page d'information d'Instagram (Huang, 2018[96]), dans lequel elle présentait les technologies utilisées par Facebook pour lutter contre le terrorisme et l'extrémisme violent en ligne, laissant ainsi penser que les deux plateformes emploient les mêmes. Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus terroristes et extrémistes violents est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. Instagram n'est pas membre du GIFCT, mais participe au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Instagram peut supprimer des contenus ou des informations partagés par les utilisateurs s'il estime qu'ils enfreignent ses Conditions d'utilisation et ses politiques (notamment ses Règles de la communauté). Il peut aussi refuser ou arrêter immédiatement de fournir tout ou partie de ses services à un utilisateur (en particulier, en résiliant ou en désactivant son compte) si celui-ci enfreint de manière claire, sérieuse ou répétée ses Conditions d'utilisation et ses politiques (notamment ses Règles de la communauté). Instagram a annoncé récemment une mise à jour de sa politique relative à la désactivation des comptes. Outre la suppression des comptes comportant un certain pourcentage de contenus en infraction (non communiqué), il supprimera aussi les comptes qui ont enregistré un certain nombre d'infractions pendant une période donnée (non communiqué non plus) (Instagram, 2019[97])
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Oui. Le rapport de transparence de Facebook de novembre 2019 comportait des informations relatives à Instagram dans quatre domaines : nudité des enfants et exploitation sexuelle, marchandises réglementées, suicide et automutilation et propagande terroriste.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	La « propagande terroriste » regroupe trois types d'informations : - *Prévalence (fréquence des violations relevant de la propagande terroriste et des contenus violents et explicites sur Instagram).* La prévalence désigne le pourcentage de vues des contenus qui enfreignaient certaines Règles de la communauté. Facebook/Instagram estime par exemple que les contenus qui

	enfreignaient ses règles relatives à la propagande terroriste ont représenté moins de 0.04 % de vues au premier trimestre 2019. Autrement dit, sur 10 000 contenus vus sur Facebook/Instagram, moins de quatre comprenaient des éléments considérés par la plateforme comme de la propagande terroriste. (Les chiffres correspondent aux décisions définitives et non aux contenus signalés comme susceptibles de constituer une violation des règles mais qui ont finalement été autorisés.) - *Contenus pour lesquels des mesures ont été prises (quantité de contenus sur lesquels Instagram est intervenu).* Les mesures prises peuvent comprendre la suppression d'un contenu, l'ajout d'un filtre d'avertissement sur des photos ou des vidéos susceptibles de déranger certains publics ou la désactivation d'un compte. Ce chiffre indique le nombre total de contenus pour lesquels Instagram a pris des mesures au cours de la période couverte par le rapport parce qu'ils ne respectaient pas les Règles de la communauté. - Taux de proactivité (taux de contenus pour lesquels Instagram a pris des mesures avant qu'ils n'aient été signalés par des utilisateurs). Ce chiffre indique le pourcentage de contenus pour lesquels Instagram a pris des mesures parce qu'ils enfreignaient ses règles avant qu'ils ne soient signalés par les utilisateurs. Il regroupe les contenus détectés grâce aux outils d'intelligence artificielle utilisés par la plateforme et par les évaluateurs humains.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	- *Prévalence.* Ce chiffre indique le nombre estimé de vues d'un contenu enfreignant les règles divisé par le nombre estimé de vues de la totalité des contenus sur Instagram sur la période couverte par le rapport. Par exemple, une prévalence de la propagande terroriste comprise entre 0.18 % et 0.20 % signifie que sur 10 000 contenus vus, 18 à 20 en moyenne ne respectaient pas les règles de la plateforme concernant la propagande terroriste. La prévalence fournit une indication de la fréquence à laquelle est vu un contenu interdit, et non le nombre total de ce type de contenus. Elle est estimée à partir d'échantillons de contenus prélevés dans différentes rubriques d'Instagram. Pour les violations relevant de la propagande terroriste, Instagram fournit uniquement une estimation de la limite supérieure, parce qu'il se dit « convaincu que la prévalence des vues des contenus enfreignant ces règles est inférieure à cette limite ». (Facebook, s.d.[71]) - *Contenus pour lesquels des mesures ont été prises.* Ce chiffre indique le nombre total de contenus pour lesquels Instagram a pris des mesures au cours de la période couverte par le rapport parce qu'ils ne respectaient pas ses règles. Instagram ne comptabilise pas les contenus qui ont été transmis aux autorités répressives. Ce chiffre regroupe les contenus pour lesquels Instagram a pris des mesures à la suite d'un signalement de la part d'un utilisateur et qui ont été détectés de manière proactive. - *Taux de proactivité.* Ce taux correspond au nombre de contenus pour lesquels Instagram a pris des mesures avant qu'ils ne

	soient signalés par les utilisateurs divisé par le nombre total de contenus pour lesquels il a pris des mesures.
10. Fréquence de publication des rapports de transparence.	Les rapports de transparence d'Instagram sont publiés avec ceux de Facebook et selon le même calendrier.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. De nombreux exemples de contenus terroristes et extrémistes violents ont été postés sur la plateforme (Carmen, 2015[98]) (Hymas, 2019[99]) (Cox, 2019[100]).

8. QQ

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation interdisent toutefois aux utilisateurs de soumettre, de mettre en ligne ou d'afficher des contenus qui, selon l'avis raisonnable de QQ ou effectivement : • enfreignent des lois ou réglementations (ou peuvent entraîner une violation des lois ou des réglementations) ; • créent un risque de perte ou de préjudice pour des personnes ; • nuisent à des personnes (adultes ou mineures) ou les exploitent de quelque façon que ce soit, notamment par des messages d'intimidation, de harcèlement ou des menaces de violence ; • sont haineux, assimilés à du harcèlement, insultants, racialement ou ethniquement offensants, diffamatoires, humiliants (publiquement ou de quelque façon que ce soit), menaçants, grossiers ou répréhensibles de quelque façon que ce soit.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://www.tencent.com/en-us/zc/termsofservice.shtml et https://www.tencent.com/en-us/zc/acceptableusepolicy.shtml[4]
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions de service ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation	QQ indique qu'il peut examiner (mais ne s'engage pas à le faire) les contenus (dont les contenus postés par ses utilisateurs) ou les programmes ou services tiers proposés sur QQ pour vérifier s'ils

ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	respectent ses politiques, les lois et réglementations en vigueur ou ne sont pas répréhensibles de quelque manière que ce soit, et se réserve le droit de bloquer ou supprimer des contenus pour quelque raison que ce soit, conformément aux lois et réglementations en vigueur.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée,
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée,
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Pas d'informations communiquées. QQ n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	QQ indique qu'il peut suspendre ou résilier l'accès à ses services s'il pense raisonnablement qu'un utilisateur a enfreint les Conditions générales d'utilisation, que l'utilisation de ses services crée un risque pour la plateforme ou d'autres utilisateurs, que la suspension ou la résiliation est requise par la législation en vigueur ou à sa seule et absolue discrétion.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.

10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

9. Youku Tudou

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Cependant, les Conditions générales d'utilisation de Youku Tudou interdisent les contenus qui incitent à la haine et à la discrimination ethnique, et/ou qui menacent l'unité ethnique, ainsi que les contenus qui incitent à commettre des délits, glorifient la violence ou prônent des activités terroristes.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page http://mapp.youku.com/service/agreement-eng
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Youku Tudou indique « gérer » les informations mises en ligne, publiées ou diffusées par les utilisateurs et prendre des mesures, telles que suspendre la diffusion, supprimer des contenus mis en ligne pour empêcher leur diffusion, conserver des enregistrements et effectuer des signalements auprès des autorités compétentes si les informations mises en ligne sont interdites par les lois et réglementations en vigueur ou enfreignent les Conditions générales d'utilisation.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et	Pas d'informations communiquées. Youku Tudou n'est pas membre du GIFCT et ne participe pas au

extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Le non-respect des Conditions générales d'utilisation peut entraîner la suppression du contenu, le blocage du contenu et des informations, la suspension, la résiliation ou l'annulation du compte de l'utilisateur ou toute autre mesure conforme aux réglementations en vigueur.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

10. QZone

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de QQ International[5] interdisent toutefois aux utilisateurs de publier, envoyer, diffuser ou enregistrer des contenus contraires à la législation ou inappropriés, insultants, obscènes ou violents.

2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://imqq.com/html/FAQ_en/html/Miscellaneous_1.html [6]
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Aucune procédure n'est indiquée.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Pas d'informations communiquées. QQ International n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales	QQ International indique qu'en cas de violation de ses Conditions générales d'utilisation, il est autorisé à suspendre la licence de l'utilisateur, interrompre ses services, appliquer des restrictions de services, reprendre le compte de l'utilisateur, mener des enquêtes légales et d'autres mesures appropriées, en fonction de la gravité du

d'utilisation ou des Règles de la communauté.	comportement de l'utilisateur, et ce sans en avertir ce dernier au préalable.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

11. TikTok

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Les Règles communautaires de TikTok précisent toutefois que « des personnes ou des organisations dangereuses » ne peuvent pas utiliser la plateforme pour promouvoir le terrorisme, le crime ou d'autres types de comportements susceptibles de porter préjudice. Les terroristes et les organisations terroristes sont expressément cités dans ce groupe. TikTok définit les terroristes et organisations terroristes comme des acteurs non-étatiques qui utilisent la violence préméditée ou la menace de violence pour causer des dommages à des civils, pour intimider ou menacer une population, un gouvernement ou une organisation internationale pour atteindre des objectifs politiques, religieux, ethniques ou idéologiques. Plus globalement, TikTok définit les personnes ou les organisations dangereuses comme les auteurs de crimes ou d'autres types de dommages graves. Ces groupes et ces crimes comprennent notamment, mais pas exclusivement, les groupes haineux, les organisations extrémistes violentes, les homicides, la traite des personnes, le trafic d'organes, le trafic d'armes, le trafic de drogue, les enlèvements, les extorsions, le chantage, le blanchiment d'argent, la fraude, la cybercriminalité.

	Les noms, symboles, logos, slogans, uniformes, gestes, portraits ou tout autre objet destiné à représenter des personnes et/ou des organisations dangereuses, ou des contenus qui font l'éloge de personnes et/ou d'organisations dangereuses, les glorifient ou les soutiennent, sont interdits sur TikTok, à l'exception des contenus pédagogiques, historiques, satiriques, artistiques ou clairement identifiés comme des contre-discours ou qui visent à sensibiliser aux dommages causés par des personnes et/ou des organisations dangereuses.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://www.tiktok.com/legal/terms-of-service?lang=fr, https://support.tiktok.com/fr/ et https://www.tiktok.com/community-guidelines?lang=fr#39
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	TikTok indique de manière générale qu'il peut à tout moment et sans notification préalable, supprimer ou désactiver l'accès à un contenu à sa seule discrétion, pour quelque raison que ce soit ou sans raison. TikTok peut supprimer un contenu qu'il estime répréhensible, contraire à ses Conditions générales d'utilisation ou ses Règles communautaires ou préjudiciable de quelque façon que ce soit à ses services ou à ses utilisateurs.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Si un utilisateur pense que son contenu a été supprimé par erreur, il peut faire appel de cette décision.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	TikTok dispose d'un dispositif qui permet aux utilisateurs de signaler les contenus qui enfreignent ses Règles communautaires. Il utilise des systèmes automatisés pour avertir son personnel de l'existence de contenus ou de comptes problématiques. Les contenus ou les comptes signalés comme potentiellement associés au terrorisme ou à l'extrémisme sont examinés par des membres spécialisés de son équipe de modération. La plateforme forme son équipe de modération aux techniques les plus récentes utilisées par les terroristes pour ne pas être détectés, dès lors qu'elles sont découvertes. Ces formations complètent

	celle reçue par tous les modérateurs pour savoir comment détecter des comptes et contenus terroristes et les distinguer des contenus et comptes qui peuvent poser problème mais sont autorisés. Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus terroristes et extrémistes violents est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. TikTok n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La violation des Règles communautaires peut entraîner la résiliation du compte et/ou la suppression du contenu.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. TikTok a toutefois publié un premier rapport de transparence en décembre 2019 dans lequel il indique les demandes légales d'informations sur les utilisateurs, les demandes de suppression de contenu de la part des autorités et les avis de retrait de contenu soumis à droit d'auteur pour le premier semestre 2019 (TikTok, 2019[25]).
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

12. Weibo

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de Weibo interdisent toutefois aux utilisateurs de mettre en ligne, afficher et diffuser des contenus offensants, insultants, intimidants, racialement discriminants, malveillants, violents ou illégaux de quelque manière que ce soit.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://www.weibo.com/signup/v5/protocol
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Weibo indique de manière générale que ses opérateurs ont le droit d'examiner, surveiller et traiter le comportement et les informations des utilisateurs de la plateforme, notamment mais pas exclusivement, les informations sur les utilisateurs (informations sur le compte, données personnelles, etc.), les données de contenu (lieu, texte, photos, audio, vidéos, marques, brevets, publications, etc.) et le comportement des utilisateurs (relations, commentaires, lettres privées, sujets et activités auxquels ils participent, données marketing, réclamations, etc.).
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données	Weibo dispose d'un dispositif qui permet aux utilisateurs de signaler des contenus illégaux ou répréhensibles. Les signalements sont ensuite vérifiés et traités par des modérateurs. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. Weibo n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.

de partage d'empreintes numériques/d'adresses URL)	
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	En cas de violation des Conditions générales d'utilisation, Weibo est autorisé à suspendre ou résilier ses services.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. La tuerie de Christchurch a été postée sur Weibo (Kenny, 2019[95]).

13. iQIYI

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation d'iQIYI interdisent toutefois la promotion du terrorisme, de l'extrémisme (pas spécifiquement l'extrémisme violent), la haine, la discrimination ethnique et la diffusion de la violence.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://www.iqiyi.com/user/register/protocol.html

3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	iQIYI indique se réserver le droit d'annuler l'accès des utilisateurs à ses produits et services ou leur capacité à créer, mettre en ligne, publier et diffuser des contenus sans avertissement préalable.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Pas d'informations communiquées. iQIYI n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	iQIYI précise qu'en cas de violation de ses Conditions générales d'utilisation, il est autorisé à suspendre ou annuler le compte concerné et de signaler le cas échéant les infractions aux autorités.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.

8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

14. Reddit

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. La Politique de contenu de Reddit interdit toutefois les contenus qui encouragent, font l'apologie, incite ou appelle à la violence ou à des préjudices physiques contre une personne ou un groupe de personnes.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://www.redditinc.com/policies/user-agreement et https://www.redditinc.com/policies/content-policy Il convient de préciser que Reddit utilise un système de modération à plusieurs niveaux. La Politique de contenu mentionnée ci-dessus encadre les contenus mis en ligne sur la plateforme, qui rassemble des milliers de communautés créées et modérées par les utilisateurs eux-mêmes, à titre volontaire. Outre la Politique de contenu, qui s'applique à l'ensemble du site, les modérateurs définissent les règles de leurs communautés, en fonction des thèmes abordés par celles-ci. Ces règles sont clairement indiquées dans un encadré affiché sur la page de chaque communauté.
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Oui. Disponibles à la pagehttps://www.redditinc.com/policies/broadcasting-content-policy. Les contenus diffusés en direct sur Reddit sont soumis à des règles supplémentaires à celles figurant dans la Politique de contenu. **Pas de contenu « inapproprié au travail » (NSFW)** Les contenus diffusés en direct ne doivent pas comprendre d'éléments « inappropriés au travail ». Comme l'indique la Politique de contenu, cette mention désigne des contenus suggestifs ou comprenant de la nudité, de la pornographie ou de la violence, qu'un

	spectateur raisonnable ne souhaite pas nécessairement voir dans un endroit public ou formel tel que son lieu de travail. **Pas de comportement illégal ou dangereux** Les contenus diffusés en direct ne doivent pas représenter des activités illégales ou qui présentent un risque déraisonnable de dommage physique pour le sujet filmé ou les personnes à proximité. **Pas de contenu soumis à quarantaine** Les contenus diffusés en direct ne doivent pas comprendre d'éléments qui seraient sinon soumis à quarantaine. Comme l'indique la Politique de contenu, ce sont des éléments que les utilisateurs moyens de la plateforme pourraient trouver très insultants ou choquants ou qui encouragent les canulars.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Sur le site, les administrateurs (des employés payés par Reddit) appliquent différentes méthodes pour faire appliquer les règles, telles que : • prier poliment l'utilisateur de retirer son contenu, • prier l'utilisateur moins poliment de retirer son contenu, • suspendre un compte de manière temporaire ou définitive, • supprimer certaines options privilégiées d'un compte ou le restreindre, • appliquer des restrictions à certaines communautés, telles que l'ajout de la mention « NSFW » ou la mise en quarantaine (voir ci-dessus), • supprimer des contenus, • interdire des communautés. Les modérateurs bénévoles disposent également d'un certain nombre de moyens pour exécuter les règles à l'échelle de leurs communautés. Ils peuvent par exemple exclure un utilisateur de la communauté (à titre temporaire ou définitif) ou supprimer ses posts. Ces mesures sont prises indépendamment des administrateurs Reddit. La mise en quarantaine (Reddit Inc., s.d.[101]) est appliquée aux communautés (qui correspondent grosso modo à des groupes partageant un intérêt commun) que les utilisateurs pourraient trouver insultantes ou choquantes ou qui visent à encourager les canulars, et qui nécessitent une surveillance supplémentaire. L'objectif de cette mesure est de faire en sorte que le contenu de ces communautés ne soit pas vu accidentellement par des personnes qui ne veulent pas le voir ou dans un contexte inapproprié. Les communautés soumises à quarantaine affichent un message d'avertissement demandant explicitement à l'utilisateur de confirmer qu'il veut bien voir ce contenu. Elles ne génèrent aucune recette, n'apparaissent pas dans les fils d'actualité accessibles sans posséder un compte Reddit (publications populaires) et ne figurent pas dans les résultats de recherches ou les recommandations. Reddit peut également appliquer d'autres restrictions à des produits existant déjà ou qui pourraient être développés à l'avenir (supprimer des outils de personnalisation, par exemple).

4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Pour qu'une communauté sorte de la quarantaine, ses modérateurs (voir la section 5, ci-après) doivent déposer un recours qui devra comprendre un compte rendu détaillé des changements apportés aux pratiques de modération de la communauté (ces changements peuvent différer selon les communautés et peuvent rassembler diverses techniques, telles que l'ajout de modérateurs, la mise en place de nouvelles règles, l'utilisation d'outils de modération plus agressifs, la modification du type de la communauté, etc.). Il devra aussi apporter des preuves de l'exécution cohérente et durable de ces changements pendant au moins un mois, afin de montrer une évolution réelle de la communauté. Reddit peut à son entière discrétion supprimer ou retirer un contenu à tout moment et pour quelque raison que ce soit, notamment en cas de violation de ses Conditions générales d'utilisation ou de sa Politique de contenu, ou si le contenu lui crée une responsabilité. Il est possible de faire appel des mesures prises par la plateforme en cas de violation de sa Politique de contenu, qu'elles concernent un compte individuel ou une communauté. Ce sont les employés de Reddit qui examinent les appels.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Reddit s'appuie sur un système de modérateurs bénévoles parmi ses utilisateurs. La modération d'une communauté est une fonction non officielle et non rémunérée. Les créateurs d'une communauté en deviennent automatiquement les premiers modérateurs et ils peuvent nommer d'autres utilisateurs pour les aider. Reddit se réserve le droit de supprimer ou de limiter la capacité d'un utilisateur à exercer la fonction de modérateur à tout moment et pour quelque raison que ce soit, notamment en cas de violation de ses Conditions générales d'utilisation. Les modérateurs doivent respecter la Politique de modération (Reddit Inc., 2017[102]), et, s'ils reçoivent un signalement concernant leur communauté, ils doivent prendre des mesures de modération en supprimant le contenu et/ou en le transmettant aux administrateurs de Reddit pour qu'ils l'examinent. Les modérateurs peuvent définir des règles pour leur communauté et les faire appliquer, à condition qu'elles ne contreviennent pas aux Conditions générales d'utilisation et aux autres politiques de Reddit. Ils peuvent paramétrer l'AutoModerator, un outil d'aide à la modération fourni par la plateforme qui permet d'effectuer automatiquement certaines tâches, telles qu'envoyer des commentaires à des posts pour rappeler les règles de la communauté aux utilisateurs ou supprimer ou étiqueter certains posts par domaine ou mot clé (Reddit Inc., s.d.[103]). Des employés Reddit formés à cet effet sont par ailleurs chargés d'exécuter la Politique de contenu sur l'ensemble de la plateforme. Enfin, les utilisateurs de Reddit participent eux-mêmes au signalement et au classement des contenus discutables. Ils peuvent

	les signaler aux modérateurs de la communauté ou aux employés de Reddit et voter contre un contenu. Un certain nombre de votes négatifs conduit à déclasser ou à masquer un contenu. Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus discutables est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. Les modérateurs bénévoles n'engendrent aucun coût pour Reddit. Reddit n'est pas membre du GIFCT, mais participe au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Le non-respect des Conditions générales d'utilisation ou de la Politique de contenu peut entraîner la suppression du contenu et/ou la suspension ou la résiliation définitive du compte de l'utilisateur (selon la gravité de l'incident), de son statut de modérateur ou de sa capacité à accéder aux services de Reddit ou à les utiliser. Les modérateurs doivent également respecter la Politique de modération sous peine d'encourir des conséquences, telles que la suppression de certaines fonctionnalités ou de certains droits associés à leur fonction de modérateur. Enfin, si une communauté ne respecte pas la Politique de contenu ou la Politique de modération, elle peut être mise en quarantaine ou interdite, selon l'ampleur ou la gravité des violations.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Pas à proprement parler, mais Reddit publie des rapports de transparence qui comprennent une partie sur les contenus supprimés en raison de la violation des règles propres à une communauté ou de la Politique de contenu, qui inclut des dispositions sur les contenus violents. Dans son dernier rapport (Reddit Inc., 2018[104]), Reddit explique que l'immense majorité des suppressions de contenus sont effectuées par les modérateurs des communautés au sein de ces dernières. Elles résultent essentiellement du non-respect des règles propres aux communautés concernées définies par celles-ci et leurs modérateurs. Même si l'exécution de ces règles et celle de la Politique de contenu de Reddit peuvent parfois se chevaucher, les mesures prises par les modérateurs restent totalement distinctes de celles des administrateurs de Reddit.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Le rapport indique le nombre de contenus supprimés par les modérateurs des communautés et les administrateurs de Reddit suite à la violation de la Politique de contenu, le nombre de signalements de non-respect de la Politique de contenu ayant entraîné ou non des mesures, et le pourcentage des violations de la Politique de contenu réparties par catégorie (sans catégorie, usurpation d'identité, données personnelles, sexualisation de mineurs, marchandises contrôlées, pornographie involontaire, contournement d'une exclusion, harcèlement, incitation à la violence et à l'auto-mutilation).
9. Méthodologies appliquées pour déterminer, calculer ou	Non communiquées.

estimer les informations et données figurant dans les rapports de transparence.	
10. Fréquence de publication des rapports de transparence.	Annuelle.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. La vidéo de l'attentat de Christchurch a été mise en ligne sur l'une des communautés de Reddit. (Hatmaker, 2019[105]) Les administrateurs de la plateforme ont exclu toute la communauté de Reddit.

15. Twitter

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des *contenus* terroristes et extrémistes violents, mais la plateforme possède une politique relative au terrorisme et à l'extrémisme violent fournissant des informations sur ce que Twitter considère comme une organisation terroriste ou extrémiste violente, ainsi que des exemples de contenu ne respectant pas ces critères. La partie Sécurité des Règles de Twitter interdit explicitement le terrorisme et l'extrémisme violent. D'après la politique relative au terrorisme et à l'extrémisme violent, les utilisateurs ne peuvent pas proférer de menaces de terrorisme ou d'extrémisme violent, ni les promouvoir. Twitter affirme qu'il n'accepte pas les organisations terroristes ou les personnes ou les groupes extrémistes violents qui leur sont associés et encouragent leurs activités illégales. Il fonde son appréciation de ces organisations sur les définitions du terrorisme national et international, sans toutefois préciser ces dernières. Les organisations sont aussi évaluées en fonction de ses propres critères de définition des groupes extrémistes. Les organisations qui : • sont identifiées par le biais de leur objectif affiché, de leurs publications ou de leurs agissements comme un groupe extrémiste, • ont commis ou commettent des actes de violence et/ou encourageant la violence pour soutenir leur cause, • visent des civils dans leurs actes et/ou pour promouvoir la violence sont considérées comme des groupes extrémistes violents. Twittter examine les activités d'un groupe sur sa propre plateforme et à l'extérieur de celle-ci pour déterminer s'il commet ou encourage des actes de violence à l'encontre de civils pour servir une cause politique, religieuse et/ou sociale. Il fournit les exemples de contenu contraires à sa politique sur le terrorisme et l'extrémisme violent suivants :

	• commettre ou promouvoir des actes au nom d'une organisation terroriste ou d'un groupe extrémiste violent, • recruter pour une organisation terroriste ou un groupe extrémiste violent, • fournir ou distribuer de services (financiers, de média/propagande) pour servir les objectifs affichés d'une organisation terroriste ou d'un groupe extrémiste violent, • utiliser les insignes ou les symboles d'organisations terroristes ou de groupes extrémistes violents pour les promouvoir. La partie des Règles sur la conduite haineuse indique que les utilisateurs ne peuvent pas menacer d'autres personnes, les harceler et inciter à la violence envers elles sur la base de critères de race, d'origine ethnique, de nationalité, d'orientation sexuelle, de sexe, d'identité sexuelle, d'appartenance religieuse, d'âge, de handicap ou de maladie grave. Les comptes qui ont pour principal objectif d'inciter à nuire en fonction de ces critères sont interdits. Les utilisateurs ne peuvent pas non plus poster d'images ou de symboles de haine pour leur bannière ou leur photo de profil, ni employer leur nom d'utilisateur, leur identifiant ou leur biographie pour se livrer à un comportement inapproprié, tel que le harcèlement ciblé ou l'envoi de messages de haines à l'encontre d'une personne, d'un groupe ou d'une catégorie précise. Cette politique interdit les menaces de violence, les messages souhaitant ou espérant un préjudice grave pour une personne ou un groupe de personne ou incitant à leur causer un préjudice grave, les références à des meurtres de masse, à des événements violents ou à des actes de violence précis dont des groupes précis ont été les principales cibles ou victimes, et les messages suscitant la méfiance envers d'un groupe spécifique. Enfin, la politique relative à l'apologie de la violence interdit l'apologie de la violence, notamment des événements violents où des personnes sont visées en raison de critères précis (tels que la race, l'origine ethnique, la nationalité, l'orientation sexuelle, le sexe, l'identité sexuelle, l'appartenance religieuse, l'âge, le handicap ou la maladie grave), ce comportement pouvant inciter à la violence motivée par la haine ou l'intolérance ou l'exacerber. En vertu de cette politique, les utilisateurs ne peuvent pas faire l'apologie de crimes violents ou d'événements violents visant des personnes précises en raison de leur appartenance à un groupe spécifique, ni les auteurs de tels actes, ni les célébrer, les louer ou les cautionner. Les messages d'« apologie » comprennent des messages louant, célébrant ou approuvant les faits, tels que « je suis heureux que cela soit arrivé », « cette personne est mon héros », « j'aimerais que plus de gens fassent cela » ou « j'espère que cela donnera envie à d'autres de faire pareil ». Les infractions à cette politique comprennent notamment mais pas exclusivement le fait de faire l'apologie, de louer, de cautionner ou de célébrer : • des actes violents commis par des civils qui ont entraîné la mort ou des blessures physiques graves, tels que des meurtres ou des tueries de masse,

	• des attentats commis par des organisations terroristes ou des groupes extrémistes violents, • des événements violents ciblant des groupes spécifiques, tels que l'holocauste ou le génocide rwandais.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://help.twitter.com/en/rules-and-policies/twitter-rules, https://help.twitter.com/en/rules-and-policies/violent-groups, https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy ethttps://help.twitter.com/en/rules-and-policies/glorification-of-violence
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Twitter possède différentes mesures d'exécution qu'il peut appliquer en cas de violation de ses Règles (Twitter, s.d.[106]). a. *Exécution au niveau d'un tweet* : s'applique au contenu qui enfreint les politiques de Twitter, mais la plateforme pense qu'il est dans l'intérêt du public que ce contenu demeure accessible. Le tweet est masqué par un avis donnant la possibilité à l'utilisateur de le voir ou non. Les tweets d'intérêt public n'apparaissent pas dans les parties Top Tweets, recherche sans risque, recommandations via les notifications push, les notifications par courriel ou par SMS, le calendrier des événements en direct et l'onglet explorer. Twitter intervient au niveau d'un tweet pour ne pas prendre des mesures trop sévères à l'encontre d'un compte habituellement correct mais qui a commis une erreur et enfreint ses règles. Les mesures prises à ce niveau peuvent consister à limiter la visibilité du tweet, exiger son retrait et le masquer jusqu'à ce qu'il soit retiré. b. *Exécution au niveau d'un message direct* : lors d'une conversation par message direct, si l'un des utilisateurs envoie un signalement concernant son interlocuteur, Twitter bloque l'envoi des messages du contrevenant à la personne qui a effectué le signalement. La conversation sera également supprimée de sa boîte de réception. S'il s'agit d'une conversation de groupe, le message direct qui a enfreint les règles peut être masqué par un avis afin qu'aucun membre du groupe ne puisse plus le voir. c. *Exécution au niveau du compte* : appliquée lorsque Twitter estime qu'un utilisateur a enfreint les règles de manière flagrante ou qu'il a continué à les enfreindre de façon répétée après

	avoir reçu des notifications de la part de Twitter. Les mesures prises peuvent être les suivantes : - Demander une modification des contenus ou du profil : si le profil ou les contenus d'un compte enfreignent ses politiques, Twitter peut le désactiver provisoirement et demander à l'utilisateur concerné de modifier les contenus ou les informations de son profil pour qu'ils respectent ses règles. Twitter précise également quelle politique ont enfreint les contenus ou le profil. - Placer un compte en mode lecture seule : si un compte habituellement correct semble enfreindre les règles, Twitter peut le placer provisoirement en mode lecture seule pour limiter sa capacité à tweeter, retweeter et aimer du contenu jusqu'à ce que la situation se calme. L'utilisateur concerné peut lire son fil d'actualité et envoyer des messages directs uniquement à ses abonnés. Les autres utilisateurs continuent à voir un compte placé en mode lecture seule et peuvent communiquer avec lui. La durée d'application de cette mesure peut aller de 12 heures à 7 jours selon la nature de l'infraction. - Vérifier la propriété du compte : pour vérifier qu'un utilisateur n'abuse pas de l'anonymat offert par Twitter pour harceler d'autres utilisateurs, la plateforme peut lui demander de confirmer la propriété du compte avec un numéro de téléphone ou une adresse électronique. Cette mesure permet d'identifier les utilisateurs qui se servent de plusieurs comptes à des fins inappropriées et d'intervenir sur ces comptes. Si un compte est bloqué en attendant la résolution d'un problème (tel que la demande d'un numéro de téléphone), il est retiré de la liste de ses abonnés, des retweets et des comptes aimés tant que le problème n'a pas été résolu (fourniture d'un numéro de téléphone, par exemple). - Suspension définitive : mesure d'exécution la plus sévère. En cas de suspension définitive, un compte n'est plus vu du tout par les autres utilisateurs et son propriétaire n'est plus autorisé à créer de nouveaux comptes.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Des notifications sont envoyées lorsque Twitter demande à un utilisateur de modifier son comportement pour respecter ses règles (demande de modification des contenus ou du profil) ou en cas de suspension définitive d'un compte. Dans ce dernier cas, Twitter informe l'utilisateur que son compte a été suspendu en raison de violations graves en indiquant quelle(s) politique(s) a ou ont été enfreinte(s) et quels contenus sont concernés.

4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Les utilisateurs peuvent faire appel des suspensions définitives s'ils estiment que Twitter a commis une erreur. Si la suspension est confirmée à l'issue de la procédure d'appel, Twitter envoie des informations relatives à la politique qui a été enfreinte par le compte.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Twitter dispose de trois moyens de détecter les contenus contraires à ses règles. 1. Le signalement par les utilisateurs Twitter encourage les utilisateurs à signaler les infractions à ses règles. Des modérateurs examinent les signalements pour déterminer si le contenu concerné enfreint effectivement des règles. Twitter possède une équipe mondiale chargée de l'exécution de ses règles 24 heures/24 dans toutes les langues acceptées sur la plateforme. 2. Les détections proactives des contenus Twitter utilise également des outils développés en interne pour repérer les infractions à ses règles, telles que la mise en ligne de contenus terroristes et extrémistes violents, à partir des contenus postés, tels que des vidéos notoirement créées par des organisations terroristes. 3. Les détections proactives des comportements Twitter utilise des outils développés en interne pour repérer les infractions à ses règles, telles que la mise en ligne de contenus terroristes et extrémistes violents, à partir des comportements des utilisateurs qui peuvent être associés à ceux d'organisations terroristes. Twitter envisage de développer sa propre technologie anti-spam pour repérer de manière proactive les contenus terroristes et extrémismes violents, les méthodes employées par certains groupes se rapprochant du spam. Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus terroristes et extrémistes violents est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. Twitter est membre du GIFCT et participe au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Des violations de la politique sur le terrorisme et l'extrémisme violent entraînent une suspension immédiate et définitive du compte concerné. Des violations de la politique sur la conduite haineuse entraînent différentes sanctions, en fonction d'un certain nombre de facteurs, tels que mais pas exclusivement, la gravité de l'infraction et les éventuelles infractions commises précédemment par l'utilisateur. Twitter peut par exemple demander à un utilisateur de supprimer le contenu contraire aux règles et placer son compte en mode lecture seule pendant un certain temps. De nouvelles violations

	entraîneront un placement du compte en mode lecture seule plus long, voire une suspension définitive du compte. Si un utilisateur fait essentiellement preuve d'un comportement inapproprié ou est réputé avoir partagé une menace violente, son compte sera suspendu définitivement dès le premier examen. Des violations de la politique relative à l'apologie de la violence entraînent des mesures différentes selon leur gravité et l'historique des violations commises par le compte. Lors de la première violation de cette politique, Twitter demande à l'utilisateur de retirer son contenu. Il bloque également de manière temporaire l'accès de l'utilisateur à son compte. Si les violations se poursuivent après l'envoi d'un avertissement, le compte est suspendu définitivement.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Oui. Les rapports de transparence de Twitter (Twitter, 2019[107]) comportent une partie sur l'exécution des Règles comprenant les politiques décrites à la section 1, ci-dessus.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Twitter indique le nombre de comptes qui ont fait l'objet d'un signalement pour de possibles violations des Règles pendant la période couverte par le rapport. Il précise le nombre de comptes pour lesquels des mesures ont été prises, répartis en six catégories correspondant aux Règles : comportement inapproprié, exploitation sexuelle d'enfants, conduite haineuse, informations privées, médias sensibles et menaces violentes. Twitter indique également le nombre de comptes suspendus en raison de violations liées à la promotion du terrorisme et le pourcentage de comptes qui a été examiné par les outils développés en interne.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Les « comptes signalés » reflètent le nombre total de comptes que les utilisateurs ont signalés comme ayant potentiellement enfreint les Règles de Twitter. Pour obtenir des chiffres fiables, Twitter ne comptabilise qu'une seule fois les comptes signalés plusieurs fois (que plusieurs utilisateurs aient signalé un compte pour la même infraction potentielle ou que plusieurs utilisateurs aient signalé le même compte pour différentes infractions potentielles). Pour ces statistiques, Twitter ne comptabilise également qu'une seule fois les signalements concernant des tweets particuliers : s'il reçoit des signalements concernant plusieurs tweets émis par le même utilisateur, il ne les comptabilise qu'une fois dans les « comptes signalés ». Les « comptes faisant l'objet de mesures » reflètent le nombre de comptes pour lesquels Twitter a pris des mesures d'exécution pendant la période couverte par le rapport. Ces mesures sont l'une de celles décrites dans la section 4, ci-dessus. Pour obtenir des chiffres fiables, Twitter ne comptabilise qu'une seule fois les comptes qui ont fait l'objet de plusieurs mesures pour la même infraction. S'il a pris des mesures sur un tweet ou un compte au

	titre de plusieurs politiques, il comptabilisera le compte pour chacune d'entre elles. Cependant, s'il a pris à plusieurs reprises des mesures pour un tweet ou un compte au titre de la même politique (il a par exemple placé provisoirement un compte un mode lecture seule, puis a demandé à l'utilisateur de modifier ses contenus ou son profil à la suite de la même violation), il comptabilisera le compte une seule fois pour la violation de cette politique.
10. Fréquence de publication des rapports de transparence.	Semestrielle.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Voir les sections 7 et 8 ci-dessus.

16. Douban

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de Douban interdisent toutefois aux utilisateurs de mettre en ligne, diffuser ou utiliser de quelque façon que ce soit des contenus contenant de la violence gratuite ou encourageant la violence, le racisme, la discrimination, l'extrémisme religieux, la haine ou les préjudices physiques de quelque nature qu'ils soient contre un groupe ou une personne, ou qui sont répréhensibles de quelque manière que ce soit.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://www.douban.com/note/732773017/
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Douban indique se réserver le droit (sans en avoir l'obligation) d'examiner à sa se ule discrétion les contenus des utilisateurs. Il précise également qu'il peut à son entière discrétion supprimer ou retirer un contenu à tout moment et pour quelque raison que ce soit, avec ou sans envoi d'une notification à l'utilisateur.

4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Pas d'informations communiquées. Douban n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	En cas de violation des Conditions générales d'utilisation, Douban est autorisé à suspendre les droits de l'utilisateur concerné à utiliser ses services ou à résilier son compte.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

17. LinkedIn

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Les Politiques de la communauté professionnelle LinkedIn comportent des parties sur le terrorisme, la violence et le contenu nuisible et choquant qui interdisent les contenus terroristes et extrémistes violents. Terrorisme « Aucune organisation terroriste ni aucun groupe extrémiste violent n'est autorisé sur notre plateforme. Nous n'acceptons pas non plus les individus affiliés à de tels groupes ou organisations pour promouvoir leurs activités. Tout contenu décrivant une activité terroriste, conçu pour recruter pour le compte d'organisations terroristes, ou qui menace de, promeut ou soutient le terrorisme de quelque manière que ce soit ne saurait être toléré. » Violence « Il est interdit de menacer de violence une personne ou un groupe, y compris de poster des déclarations d'intention de tuer ou d'infliger de graves préjudices physiques. Les individus ou les groupes ont interdiction de promouvoir ou de se livrer à de la violence, à des dommages matériels ou à toute activité de criminalité organisée. Vous ne pouvez pas utiliser LinkedIn pour exprimer votre soutien à de tels individus ou groupes ou pour glorifier tout fait de violence. » Contenu nuisible et choquant « Toute image ou autre contenu cherchant à choquer ou à humilier est interdit. Tout contenu ou activité qui promeut, organise, dépeint ou favorise les activités criminelles est interdit. Nous n'autorisons aucun contenu décrivant ou promouvant la fabrication d'armes, la toxicomanie et les menaces de vol. Le contenu ou les activités qui promeuvent ou encouragent le suicide ou tout type de blessures auto-infligées, comme l'automutilation et les troubles alimentaires, sont interdits. Si vous relevez des signes indiquant qu'une personne envisage peut-être de se faire du mal, veuillez le signaler. »
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://fr.linkedin.com/legal/professional-community-policies?
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Oui. La diffusion de contenus en direct doit respecter les Conditions générales d'utilisation et les Politiques de la communauté professionnelle et s'inscrit en outre dans un cadre très précis. Les membres du réseau qui souhaitent utiliser cette fonctionnalité doivent remplir un formulaire de candidature qui sera examiné selon un ensemble spécifique de critères. Le formulaire de candidature est disponible à cette page https://www.linkedin.com/help/linkedin/ask/lv-app LinkedIn fournit des directives et bonnes pratiques concernant la diffusion en direct, qui sont disponibles à cette page : https://www.linkedin.com/help/linkedin/answer/100225?query=linkedin%20live&hcppcid=search

4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	LinkedIn encourage ses utilisateurs à signaler les contenus qui enfreignent les Politiques de la communauté professionnelle. Lorsqu'un utilisateur signale un contenu posté par un autre utilisateur, celui-ci n'est pas informé de l'identité de l'auteur du signalement et ce dernier ne voit plus le contenu ou la conversation signalé dans son fil d'actualité ou sa boîte de réception. LinkedIn peut examiner le contenu ou la conversation signalé pour prendre des mesures supplémentaires, telles que l'envoi d'un avertissement ou la suspension du compte si le contenu est contraire à ses Conditions générales d'utilisation ou politiques.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	LinkedIn donne à ses utilisateurs la possibilité de faire appel des décisions entraînant une restriction de leurs contenus et de leurs comptes, comme le prévoient les Politiques de la communauté professionnelle. La procédure est expliquée plus précisément à cette page https://www.linkedin.com/help/linkedin/answer/82934
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes	Les utilisateurs peuvent signaler les contenus contraires aux politiques de LinkedIn. Des modérateurs examinent les signalements pour déterminer s'il convient de prendre des mesures. Microsoft, Inc., société-mère de LinkedIn, indique qu'elle supprime les contenus terroristes publiés sur les services aux consommateurs qu'elle héberge qui sont portés à sa connaissance par des outils de signalement en ligne (Microsoft, 2016[108]). Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. LinkedIn est membre du GIFCT depuis peu.

numériques/d'adresses URL)	
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La publication d'un contenu contraire aux Conditions générales d'utilisation ou politiques peut entraîner l'envoi d'un avertissement ou la suspension du compte concerné.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Pas à proprement parler. LinKedIn publie des rapports de transparence semestriels (LinkedIn, n.d.[106]) qui comportent une partie sur les demandes de suppression de contenu émanant des autorités en raison de violations de ses Conditions générales d'utilisation ou de la législation locale, ainsi qu'un rapport sur les suppressions de contenu au titre des Politiques de la communauté professionnelle. Les contenus terroristes et extrémistes violents sont indiqués dans la catégorie « violence ou contenu explicite », qui comprend les contenus exprimant une menace ou promouvant le terrorisme, la violence ou toute autre activité criminelle et les contenus extrêmement violents ou destinés à choquer ou à humilier, qui sont donc plus larges que les contenus terroristes et extrémistes violents. Le dernier rapport est disponible à la page https://about.linkedin.com/transparency/community-report
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Le nombre de contenus qui a été supprimé. LinKedIn indique aussi le nombre total de demandes de suppression de contenus émanant des autorités en raison de violations de ses Conditions générales d'utilisation ou de la législation locale, par pays, ainsi que le pourcentage des demandes qui ont entraîné des mesures de la part de LinkedIn, mais il n'existe pas de données concernant spécifiquement les suppressions de contenus terroristes et extrémistes violents.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Elles sont décrites dans le rapport de la communauté, à la page https://about.linkedin.com/transparency/community-report
10. Fréquence de publication des rapports de transparence.	Tous les six mois.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes	C'est possible. Des recherches ont montré que des extrémistes installés aux États-Unis, mais pas nécessairement des extrémistes violents, ont utilisé LinkedIn pour promouvoir leurs programmes (START (National Consortium for the Study of Terrorism and Responses to Terrorism), 2018[109]).

et extrémistes violents ?	

18. Baidu Tieba

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Cependant, les Conditions générales d'utilisation de Baidu interdisent les contenus qui incitent à la haine fondée sur la nationalité, à la discrimination ethnique, à la violence, au meurtre et au terrorisme.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://gsp0.baidu.com/5aAHeD3nKhI2p27j8IqW0jdnxx1xbK/tb/eula.html
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Aucune procédure n'est indiquée.
4.1 Notifications des suppressions ou des	Aucune notification n'est indiquée.

autres décisions d'exécution	
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Baidu Tieba possède un dispositif qui permet aux utilisateurs de signaler des contenus illégaux ou répréhensibles. Les signalements sont ensuite vérifiés et traités par des modérateurs, qui décident en dernier ressort de conserver ou de supprimer les contenus. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. Baidu Tieba n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	S'il estime qu'un utilisateur a enfreint ses Conditions générales d'utilisation, Baidu Tieba peut l'exclure à titre provisoire ou définitif, suspendre ou supprimer son compte ou lui imposer d'autres sanctions conformément aux réglementations en vigueur.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.

10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

19. Skype

1. Comment les contenus terroristes et extrémistes violents (CTEV) sont-ils définis dans les Conditions de service ou les Règles de la communauté ?	Microsoft est la société-mère de Skype. Le Contrat de services de Microsoft, qui régit Skype, interdit toute activité nuisible à d'autres personnes, telle que la publication de contenu terroriste ou extrémiste violent, des propos haineux ou des appels à la violence contre des tiers. Microsoft indique (Microsoft, 2016[108]) que, dans le cadre de ses services, un contenu terroriste désigne un contenu posté par, ou visant à soutenir, une organisation figurant sur la liste consolidée des sanctions du Conseil de sécurité des Nations Unies (Conseil de Sécurité des Nations Unies, s.d.[110]), qui représente explicitement la violence, encourage les actes violents, cautionne une organisation terroriste ou ses actes et incite à rejoindre ces groupes. La liste des sanctions des Nations Unies répertorie les groupes considérés par le Conseil de sécurité des Nations Unie comme des organisations terroristes. Il n'est pas fourni de définition de l'extrémisme violent, mais les Conditions générales de Skype interdisent aux utilisateurs de soumettre ou de publier des contenus haineux, inappropriés, illégaux, racistes, insultants ou répréhensibles de quelque manière que ce soit.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Le Contrat de services de Microsoft est disponible à la page https://www.microsoft.com/fr-fr/servicesagreement/. Voir également https://www.skype.com/fr/legal/ios/tos/#1.
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des	Skype dispose d'une procédure de notification et de retrait. S'il reçoit une notification indiquant qu'un contenu publié, téléchargé, édité, hébergé, partagé et/ou publié sur Skype (hors

Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	communications privées) est inapproprié, viole les droits d'un tiers, ou s'il souhaite supprimer ce contenu pour quelque raison que ce soit, il se réserve le droit de le supprimer automatiquement pour quelque raison que ce soit, immédiatement ou dans des délais autres qui peuvent être définis à sa seule discrétion. Ainsi que le prévoit l'Accord de services de Microsoft, « si vous enfreignez les présentes Conditions, nous pouvons (...) cesser de vous fournir les Services ou fermer votre compte Microsoft, ou bloquer l'envoi d'une communication (ex. un e-mail, un partage de fichiers ou un message instantané) vers ou depuis les Services. Nous nous réservons également le droit de supprimer ou de bloquer votre contenu des Services à tout moment si nous pensons qu'il pourrait enfreindre la réglementation applicable ou les présentes Conditions. Lors des enquêtes relatives aux infractions suspectées des présentes Conditions, Microsoft se réserve le droit de consulter votre contenu afin de résoudre le problème. »
4.1 Notifications des suppressions ou des autres décisions d'exécution	Les notifications sont envoyées à la discrétion de Microsoft. Selon le Contrat de services de Microsoft, « Si une information doit vous être communiquée concernant un service que vous utilisez, nous vous enverrons les notifications de service (...). Si vous nous avez donné votre adresse e-mail ou votre numéro de téléphone dans le cadre de votre compte Microsoft, vous êtes susceptible de recevoir des notifications de service par e-mail ou SMS, y compris pour vérifier votre identité avant d'enregistrer votre numéro de téléphone mobile et de vérifier vos achats. Vous êtes susceptible de recevoir des notifications de service par d'autres moyens (par exemple, par des messages intégrés au produit). »
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Le formulaire pour faire appel de la suspension d'un compte Microsoft est disponible à cette page https://www.microsoft.com/en-us/concern/AccountReinstatement
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Microsoft déploie toute une gamme d'outils de filtrage, de solutions d'intelligence artificielle, de partenariats extérieurs et d'opérations de modération humaine pour détecter les contenus terroristes et extrémistes violents et mener des enquêtes à ce sujet. Les utilisateurs peuvent en outre signaler les contenus contraires aux Conditions générales d'utilisation de Skype, illégaux ou répréhensibles de quelque façon que ce soit. Des modérateurs examinent les signalements pour déterminer s'il convient de prendre des mesures. Microsoft indique qu'elle supprime les contenus terroristes publiés sur les services aux consommateurs qu'elle héberge qui sont portés à sa connaissance par ses outils de signalement en ligne (Microsoft, 2016[108]). Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé.

	Skype n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques. Microsoft est toutefois un membre fondateur du GIFCT et participe au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La publication d'un contenu contraire aux Conditions générales d'utilisation ou politiques peut entraîner la résiliation ou la suspension du compte concerné et des services fournis par Skype. Voir également les sections 4 et 4.1 ci-dessus.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. Microsoft publie des rapports sur les demandes de suppression de contenu (Microsoft, 2019[111]), dont les demandes émanant des autorités qui signalent des violations de ses Conditions générales d'utilisation ou de la législation locale, mais qui ne mentionnent pas précisément les contenus terroristes et extrémistes violents. Ces rapports indiquent par ailleurs que les chiffres présentés correspondent au total des chiffres des services aux consommateurs en ligne proposés par Microsoft « par exemple, Bing, Bing Ads, OneDrive, MSN ». Ils semblent donc inclure ceux de Skype, bien que ce service ne soit pas expressément cité.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	C'est possible. Les recherches menées par le Counter Extremism Project ont montré que des individus ont consulté et diffusé des contenus de propagande extrémistes officiels sur Skype (même s'il ne s'agit pas expressément de contenus extrémistes violents) (Counter Terrorism Project, s.d.[112]).

20. Quora

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Cependant, dans la norme Soyez courtois, soyez respectueux, Quora précise au paragraphe Pas de glorification ou de promotion de la violence, qu'il peut interdire et supprimer tout contenu de tout utilisateur qui est un membre confirmé et/ou déclaré d'un groupe figurant sur la liste des Organisations étrangères considérées comme terroristes par le département

	d'État des États-Unis.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://fr.quora.com/about/tos, https://fr.quora.com/about/acceptable_use et https://fr.quora.com/Quest-ce-que-la-norme-%C2%AB-Soyez-Courtois-Soyez-Respectueux-%C2%BB
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions de service ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Quora indique qu'il a le droit mais pas l'obligation de refuser de distribuer un contenu sur sa plateforme ou de supprimer du contenu. Une violation des règles peut entraîner l'envoi d'un avertissement et si l'utilisateur persiste, il peut lui être interdit de poser des questions et d'envoyer des réponses et des commentaires (il est bloqué) ou il peut être exclu. (Quora, s.d.[113]) Les blocages et les exclusions peuvent être provisoires. Un utilisateur bloqué ou exclu peut revenir s'il décide de changer de comportement. Le blocage est généralement levé lorsque l'utilisateur demande par message privé le déblocage de son compte.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Il n'existe pas de notification de suppression, mais des avertissements en cas de contenu inapproprié, comme indiqué ci-dessus.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Si un utilisateur pense qu'il a été bloqué ou exclu par erreur, il peut faire appel de cette décision.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les contenus qu'ils estiment contraires aux règles de Quora. Les signalements sont transmis à l'équipe de modération de Quora pour examen. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. Quora n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions	Les contenus contraires à la norme Soyez courtois, soyez respectueux peuvent être signalés aux administrateurs, qui

générales d'utilisation ou des Règles de la communauté.	peuvent ensuite les supprimer, et les infractions à cette norme peuvent entraîner l'envoi d'un avertissement, le blocage des commentaires, le blocage ou l'exclusion de l'utilisateur concerné (voir la section 4, ci-dessus). Selon la gravité de l'infraction, l'utilisateur peut être exclu immédiatement (sans avertissement ni blocage préalables). Quora peut aussi résilier ou suspendre le compte d'un utilisateur qui a enfreint l'une de ses règles.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Des questions sur la façon de rejoindre une organisation terroriste ont été posées sur Quora (Lange, 2017[114]).

21. Snapchat

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Cependant, dans les Règles communautaires de Snapchat, Snap indique dans la partie Terrorisme qu'il est interdit aux organisations terroristes et aux groupes haineux d'utiliser sa plateforme et qu'il n'a aucune tolérance pour les contenus qui prônent ou font la promotion de l'extrémisme violent ou du terrorisme. Le terme « organisations terroristes » n'est pas défini. Snap interdit également tout contenu qui encourage la discrimination ou la violence sur la base de la race, de l'appartenance ethnique, de l'origine nationale, de la religion, de l'orientation sexuelle, de l'identité de genre, du handicap ou du

	statut d'ancien combattant.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://snap.com/fr-FR/terms et https://snap.com/fr-FR/community-guidelines
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Sans objet. Snapchat ne propose pas la diffusion de contenu en direct.
4. Politiques et procédures d'application et d'exécution des Conditions de service ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Snap indique qu'il se réserve le droit de supprimer tout contenu (i) qui enfreint selon lui ses Conditions générales d'utilisation ou ses Règles communautaires, ou (ii) si nécessaire pour satisfaire ses obligations légales. Snap précise qu'il soutient les Santa Clara Principles on Transparency and Accountability in Content Moderation (Santa Clara University's High Tech Law Institute, s.d.[115]), selon lesquels les entreprises doivent informer les utilisateurs dont le contenu est supprimé ou dont le compte est suspendu des raisons de cette suppression ou de cette suspension. Les Principes disposent également que les entreprises doivent donner la possibilité de faire appel des suppressions de contenu et des suspensions de compte. Les règles de Snapchat ne prévoient toutefois pas pour l'instant l'envoi de notifications en cas de suppression de contenu ni de procédures d'appel des décisions de suppression de contenu ou de suspension de compte.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs,	Les utilisateurs peuvent signaler les contenus contraires aux règles de Snapchat (Snap Inc., s.d.[116]). Snap dispose d'une équipe spécialisée dans la sécurité qui travaille 24 heures/24. Les contenus qui enfreignent les règles

évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	de Snapchat sont supprimés. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. Snapchat n'est pas membre du GIFCT, mais participe au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Si un utilisateur enfreint ses Conditions générales d'utilisation ou ses Règles communautaires, Snapchat peut supprimer le contenu inapproprié, supprimer le compte concerné et avertir les autorités. Si un compte est supprimé parce qu'il a enfreint les règles de Snapchat, son propriétaire ne peut plus utiliser Snapchat.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. Snapchat publie des rapports de transparence (Snap Inc., 2019[117]) qui comportent une partie sur les demandes de suppression de contenu émanant des autorités qui signalent des violations de ses Conditions générales d'utilisation ou de la législation locale, mais qui ne mentionnent pas précisément les contenus terroristes et extrémistes violents.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Par exemple, la vidéo de l'attentat terroriste de Nice, en 2016, a été diffusée dans les stories de Snapchat et apparaissait dans les recherches effectuées avec la fonction Explorer.

22. Viber

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. La Politique de contenu public de Viber indique toutefois que les contenus excessivement violents, notamment ceux qui font l'apologie ou incitent à la violence, sont interdits. Ces contenus recouvrent les représentations ou les descriptions de la violence et les menaces crédibles de violence à l'encontre d'une personne et/ou d'un groupe. Viber interdit la planification ou la promotion d'actes violents qui pourraient causer directement ou indirectement des préjudices physiques ou psychologiques aux autres.

Règles de la communauté ?	
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://www.viber.com/fr/terms/viber-terms-use/ et https://www.viber.com/terms/viber-public-content-policy/
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Sans objet. Pour l'instant, Viber ne propose pas la fonction de diffusion de contenu en direct.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Viber indique que s'il approuve un compte public, l'utilisateur devient automatiquement administrateur de compte public et administrateur de dialogue en ligne public. De même, s'il crée une communauté, un utilisateur en devient automatiquement le « superadmin ». Les administrateurs doivent veiller à ce que les contenus envoyés et affichés dans leur compte public ou leur communauté respectent les politiques et Conditions générales d'utilisation de Viber, ainsi que toutes les lois et réglementations en vigueur. Ils ne doivent pas avoir un comportement interdit par l'un de ces textes ni autoriser un tiers à le faire. Viber peut supprimer tout ou partie des contenus s'il estime qu'ils sont interdits, illégaux ou contraires à ses politiques.

4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les contenus contraires aux politiques de Viber. Viber examine les signalements pour décider des mesures les plus appropriées. Les administrateurs ont la possibilité de supprimer les contenus ne respectant pas les politiques de leurs comptes et communautés. Il est difficile de déterminer dans quelle mesure les contenus publiés Viber sont modérés. Ses Conditions générales d'utilisation disposent qu'il ne s'engage pas à surveiller les dialogues en ligne publics ou les forums et que les contenus qui y sont postés ne relèvent pas de sa responsabilité. En outre, ses principales fonctionnalités étant cryptées, il s'avère impossible de modérer les contenus qui sont diffusés sur celles-ci. Les fonctionnalités publiques, telles que les communautés ou les dialogues en ligne, ne sont toutefois pas cryptées de bout en bout et Viber peut en examiner les contenus à la suite d'un signalement et les supprimer le cas échéant. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. Les utilisateurs modérateurs n'engendrent aucun coût pour Viber. Viber n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Les contenus qui enfreignent les politiques de Viber ou qui sont jugés répréhensibles par la plateforme sont supprimés. Dans ce cas, Viber peut suspendre ou résilier les comptes d'un utilisateur et bloquer ce dernier sur les dialogues en ligne publics.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.

8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. L'EIIL a créé un compte Nashir News Agency (agence de diffusion dans les médias associée à l'EIIL) sur Viber (Site Intelligence Group Enterprise, 2018[118]), (Katz, 2019[119]) qui l'a clôturé dès qu'il l'a découvert.

23. Pinterest

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Le Guide pour la communauté de Pinterest interdit toutefois les images montrant l'usage de la violence ou glorifiant la violence, ainsi que les contenus utilisés pour menacer ou organiser de la violence ou soutenir des organisations violentes. Le terme « organisations violentes » n'est pas défini. Par ailleurs, Pinterest interdit tout contenu présentant un risque réel de préjudice physique ou matériel, représentant une menace, organisant de la violence ou incitant autrui à la violence. Pinterest exclut toute personne ou groupe dont l'objectif est de porter préjudice à autrui. Le terme « organisations terroristes » n'est pas défini. Le terme « organisations terroristes » n'est pas défini.
2. Manière dont les Conditions générales d'utilisation ou les	Disponibles aux pages https://policy.pinterest.com/fr/terms-of-service et https://policy.pinterest.com/fr/community-guidelines

Règles de la communauté sont communiquées.	
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Sans objet. Pinterest ne propose pas la diffusion de contenu en direct.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Pinterest indique se réserver le droit de supprimer ou de modifier les contenus des utilisateurs ou de modifier la manière dont ils sont utilisés sur la plateforme, pour quelque raison que ce soit. Cela concerne aussi les contenus des utilisateurs qui ne respectent pas ses règles. Le Guide pour la communauté indique que Pinterest supprime certains types de contenu et que d'autres sont retirés des espaces publics de la plateforme, sans fournir davantage de détails.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Pinterest avertit « dans la plupart des cas » les utilisateurs de la suppression de leur contenu, mais il n'est pas précisé dans quels cas des notifications sont effectivement envoyées.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Il n'existe pas de mécanisme de recours en cas de suppression de contenu, mais il est possible de faire appel des suspensions de compte (Pinterest, s.d.[120]).
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Pinterest possède un dispositif qui permet aux utilisateurs de signaler les contenus qui enfreignent ses règles. Il dispose d'une équipe de modérateurs qui surveillent les contenus. Les contenus terroristes et violents sont supprimés lorsqu'ils sont détectés. Pinterest indique qu'il collabore avec les entreprises du secteur, les autorités et les experts en sécurité pour détecter les groupes terroristes. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. Pinterest est membre du GIFCT, mais ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	En cas de violation de ses règles, Pinterest peut résilier ou suspendre l'accès de l'utilisateur concerné à la plateforme immédiatement et sans avertissement préalable. Les notifications relatives à ces mesures sont envoyées à la discrétion de Pinterest.

7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. Pinterest publie des rapports de transparence (Pinterest, 2019[121]) qui comportent une partie sur les demandes de suppression de contenu émanant des autorités et de tiers privés signalant des violations de ses Conditions générales d'utilisation ou de la législation locale, mais qui ne mentionnent pas précisément les contenus terroristes et extrémistes violents.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

24. Vimeo

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Vimeo interdit toutefois les contenus qui encouragent ou soutiennent des « groupes haineux ou de terreur », représentent des actes illégaux ou de violence extrême ou fournissent des instructions sur la manière de fabriquer des engins explosifs/incendiaires ou des armes artisanales/improvisées. Les membres d'un « groupe haineux ou de terreur » ne peuvent pas créer un compte Vimeo (Cheah, 2019[122]). Le terme « groupes haineux ou de terreur » n'est pas défini.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://vimeo.com/terms et https://vimeo.com/help/guidelines
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des	Vimeo indique que le contexte joue un rôle essentiel dans l'application de ses règles et de ses processus. Si des contenus

Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	interdits apparaissent dans le contexte d'une story ou dans la description d'une œuvre dramatique, ils seront probablement conservés, mais si l'œuvre mise en ligne vise principalement à diffuser un point de vue expressément interdit par Vimeo, elle sera supprimée. Vimeo tient également compte des discours tenus par l'utilisateur concerné sur d'autres plateformes (telles que les réseaux sociaux, des blogs ou d'autres espaces où sont clairement exposées les opinions personnelles) en examinant ses intentions et sa bonne foi (Cheah, 2019[122]). En règle générale, les modérateurs de Vimeo suppriment les vidéos qui montrent des scènes de meurtres, de tortures ou d'agressions physiques ou sexuelles ou qui présentent des images choquantes, macabres ou susceptibles d'inspirer du dégoût. Vimeo comprend néanmoins que des vidéos traitent de ces sujets d'une manière réfléchie et critique. Les vidéos qui filment le monde réel comprennent parfois des scènes violentes. Le contexte est donc essentiel et les vidéos documentaires ou journalistiques disposent d'une plus grande latitude pour montrer des scènes de violence ou leurs répercussions. Pour ne pas être supprimées, ces vidéos ne doivent pas comporter d'images à sensation, gratuites ou d'exploitation. Elles doivent aussi porter la mention « adulte ». Les vidéos qui recrutent pour des organisations terroristes ou font leur propagande, qu'elles montrent ou non des scènes de violence, ne sont jamais autorisées (Vimeo, s.d.[123]).
4.1 Notifications des suppressions ou des autres décisions d'exécution	Certaines décisions de suppression de contenu font l'objet d'une notification, telles que celles concernant les violations de droits d'auteur. Vimeo n'informe toutefois pas les utilisateurs concernés de la suppression d'une vidéo ou d'un compte (et ne propose pas de mécanisme de recours) si les contenus supprimés relèvent de certaines catégories, telles que la suspicion de maltraitance d'enfants ou le terrorisme.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Il est possible de faire appel des suppressions pour violation de droits d'auteur, mais il n'existe pas de mécanisme de recours en cas de suppression de contenu terroriste et extrémiste violent.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les contenus contraires aux politiques et aux règles de Vimeo. Vimeo indique qu'il peut surveiller les comptes, les contenus et le comportement des utilisateurs, quels que soient leurs paramètres de confidentialité. Vimeo a conclu un accord avec Active Fence pour participer à la détection des contenus terroristes et extrémistes violents et devrait mettre en œuvre ce partenariat début 2020. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est

	probablement relativement élevé. Vimeo n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	En cas de violation de ses règles et Conditions générales d'utilisation, Vimeo peut, à sa seule discrétion, résilier, suspendre ou limiter l'accès de l'utilisateur concerné à son compte ou à ses contenus, et clôturer le compte concerné.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

25. IMO

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation d'IMO interdisent toutefois l'utilisation de ses services pour diffuser des menaces de violence.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://imo.im/policies/terms_of_service et https://imo.im/policies/acceptable_use_policy.html
3. Existe-t-il des dispositions précises applicables aux	Non.

contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	IMO indique se réserver le droit de supprimer, filtrer, modifier ou désactiver l'accès à des contenus sans en avertir le propriétaire au préalable, s'il considère, à sa seule discrétion, qu'ils enfreignent ses règles ou nuisent de quelque manière que ce soit à ses services.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	IMO indique qu'il n'est soumis à aucune obligation d'examiner les contenus, mais qu'il se réserve le droit de le faire à quelque moment que ce soit. Il ne précise toutefois pas quels types d'examens sont réalisés. IMO n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La violation des règles d'IMO peut entraîner la suspension ou la clôture du compte de l'utilisateur concerné.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.

8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

26. Telegram

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de Telegram interdisent toutefois la promotion de la violence sur ses chaînes publiques. Cette interdiction ne s'applique pas aux « dialogues en ligne secrets ».
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://telegram.org/tos
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Aucune procédure n'est indiquée. Telegram dispose que s'il reçoit une ordonnance judiciaire confirmant qu'un utilisateur est suspecté de terrorisme, il peut communiquer l'adresse IP et le numéro de téléphone de cet utilisateur aux autorités concernées. Il précise que cela ne s'est encore jamais produit (Telegram, s.d.[124]).

4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Telegram autorise les utilisateurs à signaler les contenus qui ne respectent pas ses règles. Il dispose aussi d'une équipe qui examine les contenus des chaînes publiques. Telegram a créé en 2016 la chaîne « ISIS Watch » pour montrer les efforts qu'il déploie pour supprimer les contenus encourageant le terrorisme sur ses chaînes et bots publics. D'après la chaîne, Telegram aurait supprimé plus de 200 000 chaînes et bots publics en lien avec l'EIIL (Telegram, s.d.[125]). Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus problématiques est probablement relativement élevé. Telegram n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Aucune notification n'est indiquée.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Plusieurs attentats terroristes ont été coordonnés sur Telegram (Bennett, 2019[126]) (Hayden, 2019[127]) (Bennett, 2019[126]) (Hayden, 2019[127]).

27. LINE

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de LINE interdisent toutefois la publication ou la diffusion de contenus violents. Les activités qui bénéficient à des groupes anti-sociaux ou qui collaborent avec ces derniers sont également interdites. Le terme « groupes anti-sociaux » n'est pas défini.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://terms.line.me/line_terms/
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Oui, elles sont disponibles à la page https://terms2.line.me/LINELIVE_ToC_ME1
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	LINE applique un processus composé de deux étapes pour surveiller les posts mis en ligne sur son fil d'actualité, sur LINE LIVE, LINE Manga, LINE Fortune, LINE Pasha, LINE Step, LINE BLOG, LINE Delima et WizBall. Le contenu publié par un utilisateur sur l'un des services LINE est tout d'abord contrôlé par le système de surveillance automatique de la plateforme pour vérifier qu'il ne comprend pas de mots interdits, qu'il n'enfreint aucune règle de la plateforme et respecte ses Conditions générales d'utilisation ainsi que la législation en vigueur. Si un contenu répréhensible est détecté par le système de surveillance, il est suspendu dès sa mise en ligne. Une équipe de surveillance vérifie ensuite les contenus que le système automatique ne peut pas traiter. Elle les évalue en fonction d'une série de critères et les compare à des exemples de contenus rencontrés précédemment pour déterminer s'ils sont autorisés ou non. Si elle décide qu'ils sont contraires aux Conditions générales d'utilisation de LINE ou à la législation en vigueur, ils sont suspendus (LINE, 2019[128]). LINE ne peut pas surveiller les messages envoyés ou reçus par un utilisateur dans un salon de discussion LINE classique, sauf si l'utilisateur lui communique les données non cryptées de la conversation en utilisant l'outil de signalement (LINE, 2019[128]).
4.1 Notifications des suppressions ou des autres décisions d'exécution	Il n'existe pas de notification de suppression.

4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Les utilisateurs peuvent faire appel des décisions de suppression par l'intermédiaire du formulaire de contact de LINE.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les contenus contraires aux règles de LINE. Les signalements sont examinés par l'équipe de LINE qui « prend des mesures appropriées » (LINE, s.d.[129]) si elle constate une violation de ces règles. En plus de traiter les signalements effectués par les utilisateurs, l'équipe et le système automatique de surveillance de LINE examinent avec sérieux les posts des utilisateurs (voir la section 4, ci-dessus). Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus terroristes et extrémistes violents est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. LINE n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	LINE peut supprimer les contenus ou suspendre ou supprimer un compte sans avertissement préalable, s'il estime que l'utilisateur enfreint ou a enfreint ses règles.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. LINE publie néanmoins des rapports de transparence sur trois sujets : les demandes de communication des données utilisateurs et/ou de suppression émanant des autorités répressives, les mesures prises à l'encontre des posts qui enfreignent ses Conditions générales d'utilisation ou la législation en vigueur et l'état de déploiement du cryptage des messages et des appels (LINE, 2019[128]).
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Le rapport sur les mesures prises à l'encontre des posts en infraction sur ses services indique le nombre de contenus suspendus et leur pourcentage par catégorie, à savoir le spam, les contenus obscènes, les sollicitations, l'utilisation commerciale non autorisée des comptes, les contenus gênants et problématiques, la promotion des activités illégales et la catégorie « autres ». Les contenus terroristes et extrémistes violents semblent relever de la catégorie « promotion des activités illégales » (d'après les exemples cités dans la section 9, ci-après), mais cela n'est pas indiqué de manière explicite.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données	LINE précise que les contenus gênants et problématiques peuvent être « des remarques excessivement haineuses, des photos de cadavres, de la fraude au clic, des liens vers des sites d'hameçonnage, etc. » et que les activités illégales peuvent

figurant dans les rapports de transparence.	recouvrir « des annonces d'attaques ou d'attentats, la vente de stupéfiants, la vente de données en ligne (comptes, monnaies, avatars) contre de l'argent, etc. ».
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

28. Ask.fm

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Règles de la communauté d'Ask.fm indiquent toutefois que les organisations terroristes et les groupes extrémistes violents qui cherchent à encourager ou à commettre des activités terroristes ou criminelles violentes ne sont pas autorisés à être présents sur la plateforme pour promouvoir leurs campagnes ou leurs plans, célébrer leurs actes de violence, lever des fonds ou recruter des jeunes. Les termes « organisations terroristes » et « groupes extrémistes violents » ne sont pas définis. Par ailleurs, les utilisateurs ne peuvent pas publier de contenus comportant des menaces de quelque nature qu'elles soient, telles que des menaces de violence physique envers eux-mêmes ou autrui, ou incitant à commettre des actes de violence envers eux-mêmes ou autrui. Les mots « terroriste », « terrorisme » ou « extrémisme » ne sont pas explicitement définis.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://about.ask.fm/legal/2019-07/en/terms.html et https://about.ask.fm/community-guidelines/
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Sans objet. Ask.fm ne propose pas de diffusion de contenu en ligne.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les	Ask.fm indique se réserver le droit de surveiller l'accès des utilisateurs à ses services ou leur utilisation pour vérifier qu'ils respectent ses Conditions générales d'utilisation et d'examiner ou de modifier les contenus. Il dispose également qu'il peut bloquer ou désactiver l'accès à des contenus considérés comme répréhensibles ou nuisibles pour autrui sans avertissement préalable.

autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	
4.1 Notifications des suppressions ou des autres décisions d'exécution	Les contenus qui ne respectent pas les Conditions générales d'utilisation ou les Règles de la communauté sont supprimés. L'utilisateur concerné peut alors recevoir un avertissement, mais il n'est pas indiqué dans quel cas précisément.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Les utilisateurs dont le compte a été supprimé peuvent faire appel de cette décision.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les contenus qu'ils estiment contraires aux règles d'Ask.fm. Les signalements sont transmis à l'équipe d'Ask.fm pour examen. Ask.fm affirme examiner tous les signalements et indique également qu'il peut consulter les informations et les contenus d'un utilisateur s'il estime raisonnablement nécessaire d'exécuter ses Conditions générales d'utilisation et de protéger la sécurité de ses utilisateurs ou du public. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. Ask.fm n'est pas membre du GIFCT, mais participe au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La violation des Conditions générales d'utilisation peut entraîner la suspension ou la clôture du compte de l'utilisateur concerné ou la résiliation de son accès aux services.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.

11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Un compte Ask.fm aurait par exemple fourni des conseils sur la manière de rejoindre les combattants de l'EIIL en Iraq, ainsi que des informations sur les armes fournies à l'arrivée. (Miller, 2014[130])

29. Twitch

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Lignes de conduite de la communauté de Twitch disposent toutefois que les actes et menaces de violence seront pris au sérieux et sont considérés comme des violations qui seront traitées avec une tolérance zéro. Tous les comptes associés à de telles activités seront suspendus pour une durée indéterminée. Cela inclut notamment mais pas exclusivement : • les tentatives ou menaces d'attaque physique ou de meurtre à l'encontre d'autres personnes, • l'utilisation d'armes pour menacer, intimider, blesser physiquement ou tuer d'autres personnes. Twitch interdit également les comportements haineux, qui désignent tout contenu ou toute activité qui promeut ou encourage la violence sur la base en particulier de la race, de l'ethnie, de l'origine nationale, de la religion, du sexe, du genre, de l'identité sexuelle, de l'orientation sexuelle, du handicap, de l'état de santé et du statut d'ancien combattant.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://www.twitch.tv/p/fr-fr/legal/community-guidelines/, https://www.twitch.tv/p/fr-fr/legal/terms-of-service/ et https://help.twitch.tv/s/article/about-account-suspensions-dmca-suspensions-and-chat-bans?language=fr
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Twitch prend des mesures d'exécution à l'encontre des comptes qui enfreignent ses Conditions générales d'utilisation ou ses Lignes de conduite de la communauté. Il prend en compte plusieurs facteurs lors de l'examen des signalements d'infraction, notamment l'intention et le contexte, les dommages potentiels occasionnés à la communauté, les obligations légales et d'autres éléments. Selon la nature de l'infraction, Twitch applique différentes mesures qui comprennent l'envoi d'un avertissement, la suspension provisoire d'un compte ou sa suspension définitive pour les infractions les plus graves. Un avertissement est un préavis. Twitch peut aussi retirer les contenus associés à l'infraction. La répétition d'une infraction pour laquelle l'utilisateur a déjà reçu un avertissement, ou la

	commission d'une infraction semblable, entraînera une suspension. Les suspensions temporaires durent de 24 heures à des périodes pouvant dépasser 30 jours. En cas de suspension, l'utilisateur ne peut plus consulter ni utiliser les services de Twitch, notamment regarder ou diffuser des vidéos en direct et converser en ligne. Il peut de nouveau utiliser les services de Twitch à l'issue de la suspension. Twitch archive les infractions et plusieurs suspensions consécutives peuvent entraîner une suspension définitive. Pour les infractions les plus graves, Twitch suspend le compte immédiatement et définitivement, sans possibilité d'appel.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Des avertissements sont envoyés, selon la nature de l'infraction.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Dans les cas d'infractions n'entraînant pas une suspension immédiate, si l'utilisateur pense qu'il n'a pas enfreint les Lignes de conduite de la communauté, il peut faire appel de la décision d'exécution. Il doit indiquer dans son recours la raison pour laquelle il estime que la décision n'est pas justifiée. Lorsque l'appel a été examiné, Twitch informe l'utilisateur du résultat.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Twitch met à la disposition des utilisateurs des outils leur permettant de signaler des contenus ou des comportements qui enfreignent les Lignes de conduite de la communauté, que ce soit dans les vidéos diffusées en direct, dans le dialogue en ligne ou dans des commentaires associés à une vidéo. Les signalements sont examinés par l'équipe de sécurité de Twitch, ceux faisant état de violence extrême et de terrorisme étant traités en priorité. Un deuxième niveau de modération repose sur une série d'outils qui permettent au propriétaire d'une chaîne (parfois appelé « diffuseur ») de nommer d'autres utilisateurs modérateurs de la chaîne. Ces derniers sont alors en mesure d'exclure les utilisateurs irrespectueux, de supprimer des messages du dialogue en ligne et de prendre les mêmes mesures que le propriétaire de la chaîne. Enfin, Twitch propose aux propriétaires de chaînes un outil utilisant des algorithmes d'apprentissage automatique et de traitement du langage naturel pour empêcher l'affichage des messages à la vue des autres utilisateurs du dialogue en ligne tant qu'ils n'ont pas été examinés par un modérateur de la chaîne. Cette méthode est appelée « AutoMod » (Twitch, s.d.[131]). Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus discutables est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. Les utilisateurs modérateurs n'engendrent aucun coût pour Twitch.

	Twitch appartient à Amazon, qui a rejoint le GIFCT en septembre 2019.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La violation des Lignes de conduite de la communauté peut entraîner la suppression des contenus, l'envoi d'un avertissement et/ou la suspension du compte. Les infractions graves sont sanctionnées par une suspension immédiate du compte.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. Amazon, la société-mère de Twitch publie des rapports de transparence (Amazon, s.d.[132]), qui ne comprennent toutefois aucune information sur les contenus terroristes et extrémistes violents.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Lors d'une attaque coordonnée sur Twitch en mai 2019, des utilisateurs ont diffusé des contenus choquants, notamment des extraits de la vidéo de l'attentat de Christchurch. (Marshall, 2019[133]) Plus récemment, l'auteur de l'attentat de Halle, en Allemagne, a diffusé son attaque en direct sur Twitch. (British Broadcasting Corporation (BBC), 2019[134]) Elle a été vue par environ 2 500 personnes avant que Twitch ne retire la vidéo, qui n'a ensuite pas réapparu.

30. Xigua

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de Xigua interdisent toutefois la promotion du terrorisme et de l'extrémisme.
2. Manière dont les Conditions générales d'utilisation ou les	Disponibles à la page https://www.ixigua.com/user_agreement/

Règles de la communauté sont communiquées.	
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Aucune procédure n'est indiquée.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les activités ou les contenus illégaux. Xigua n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La violation des Conditions générales d'utilisation peut entraîner la clôture du compte de l'utilisateur concerné ou la résiliation de son accès aux services de Xigua sans avertissement préalable.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.

8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

32. Tumblr

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Les Règles communautaires de Tumblr indiquent toutefois que les contenus qui promeuvent, incitent ou encouragent les actes terroristes ne sont pas tolérés. Cela inclut les contenus qui soutiennent ou glorifient des organisations terroristes, leurs leaders ou les activités violentes qui y sont liées. Le terme « organisations terroristes » n'est pas défini.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://www.tumblr.com/policy/fr/terms-of-service et https://www.tumblr.com/policy/fr/community
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation	Si Tumblr conclut qu'un utilisateur a enfreint ses règles, il lui envoie un avertissement par courrier électronique. Si l'utilisateur refuse de se justifier ou de corriger son comportement, Tumblr peut prendre des mesures contre son compte. La plateforme précise qu'elle se

ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	réserve le droit de suspendre un compte ou de supprimer du contenu, sans préavis, quelle que soit la raison, pour protéger ses services, ses infrastructures, ses utilisateurs ou sa communauté.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Il n'existe pas de notification de suppression.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Les utilisateurs peuvent contacter le Centre d'aide de Tumblr pour faire appel d'une décision de suppression de contenu.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les activités ou les contenus illégaux. Tumblr indique que ses spécialistes formés examinent les contenus signalés et prennent des « mesures appropriées ». Les signalements n'entraînent pas toujours la suppression du contenu, les équipes spécialisées de Tumblr estimant parfois que le contenu signalé n'enfreint pas les Règles communautaires. Outre les signalements envoyés par les utilisateurs, Tumblr utilise des outils automatisés pour détecter les contenus pouvant être associés à du terrorisme ou de l'extrémisme violent et les soumettre à un examen humain. Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus discutables est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains est probablement relativement élevé. Tumblr n'est pas membre du GIFCT, mais participe au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Tumblr peut résilier ou suspendre l'accès de l'auteur de l'infraction aux services ou sa capacité à les utiliser, immédiatement et sans avertissement préalable ou responsabilité.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	No. Oath, qui détenait Tumblr auparavant (Alexander, 2019[135]), publie des rapports de transparence, qui portaient également sur Tumblr jusqu'en 2018. Ils sont toutefois très larges et ne répartissent pas les informations entre les différentes sociétés contrôlées par Oath (les demandes de suppression émanant des autorités concernent par exemple à la fois Yahoo et Tumblr). Ils ne comprennent pas non plus d'informations particulières sur les contenus terroristes et extrémistes violents (Verizon Media, 2019[136]).

8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Tumblr comporterait un grand nombre de pages faisant l'apologie du nazisme, du suprémacisme blanc, du nationalisme ethnique et du terrorisme d'extrême droite.

33. Flickr

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Les Règles de la communauté de Flickr interdisent toutefois la publication de contenus associés au terrorisme.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://www.flickr.com/help/terms et https://www.flickr.com/help/guidelines
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Flickr s'appuie sur des utilisateurs modérateurs pour détecter les contenus comprenant des éléments de nudité ou indécents, mais ce système ne s'applique pas aux contenus terroristes et extrémistes violents dans la mesure où la publication d'un contenu de cette nature entraîne la suppression du compte concerné. Les critères de détection des contenus terroristes et extrémistes violents ne sont toutefois pas précisés.

4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les contenus qu'ils estiment contraires aux Règles de la communauté. Le personnel de Flickr examine les signalements pour déterminer si le contenu enfreint effectivement les règles et prend des mesures le cas échéant. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. Flickr n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La publication de contenus terroristes et extrémistes violents entraîne la suppression du compte concerné. Flickr indique qu'il peut signaler ces comportements aux autorités répressives.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Un monument virtuel a été créé sur la plateforme pour les combattants du djihad tués en Syrie, comprenant leur nom et leur origine, ainsi que des commentaires louant leur dévotion et leur force de combat (Weimann, 2014[137]).

34. Huoshan

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de Huoshan interdisent toutefois la promotion du terrorisme et de l'extrémisme (pas spécifiquement l'extrémisme violent).
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://www.huoshanzhibo.com/agreement/
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Aucune procédure n'est indiquée. Huoshan indique qu'il tient un registre des infractions présumées à la législation et aux réglementations et des crimes présumés, qu'il signale ces cas aux autorités compétentes conformément à la loi et qu'il coopère à toutes les enquêtes relatives à ces cas.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les activités ou les contenus illégaux. L'équipe de modérateurs de Huoshan examine les signalements et prend des mesures le cas échéant. Huoshan possède aussi une équipe spécialisée dans la modération de contenu et accentue ses efforts pour améliorer ses « normes de contrôle » (Yoo, 2018[138]). Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé.

	Huoshan n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	En cas de violation de ses Conditions générales d'utilisation, Huoshan peut supprimer les posts ou les commentaires concernés, restreindre tout ou partie des fonctions du compte concerné ou résilier son accès aux services.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

35. VK

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de VK interdisent toutefois aux utilisateurs de mettre en ligne, conserver, publier, diffuser, mettre à disposition ou utiliser de quelque manière que ce soit des informations qui contiennent des éléments extrémistes et qui encouragent des activités criminelles ou qui fournissent des conseils, des instructions ou des guides relatifs à des activités criminelles. VK reprend la définition de « contenus terroristes » figurant dans la législation russe.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://vk.com/terms et https://vk.com/licence

3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Aucune procédure n'est indiquée. VK indique qu'il se réserve le droit, à sa seule discrétion et la suite de la réception d'informations de la part d'utilisateurs ou de tiers, de modifier (modérer), bloquer ou supprimer tout contenu publié enfreignant ses Conditions générales d'utilisation, ou de suspendre, limiter ou résilier l'accès de l'auteur de l'infraction à tout ou partie de ses services à tout moment, avec ou sans avertissement préalable. VK se réserve également le droit de supprimer la page d'un utilisateur et/ou de suspendre, limiter ou résilier l'accès de l'utilisateur à ses services s'il pense que l'utilisateur représente une menace pour la plateforme et/ou pour les utilisateurs.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Les suppressions de contenus sont notifiées aux utilisateurs, y compris celles concernant les contenus figurant sur la liste fédérale des contenus extrémistes du ministère de la Justice de la Fédération de Russie.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Si un utilisateur conteste la suppression ou le blocage d'un contenu, il peut contacter le Centre d'aide de VK.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	VK utilise une méthode de modération hybride. Il répond aux signalements envoyés par les utilisateurs, les agences de réglementation et les autres organismes, et effectue également une surveillance en interne avec des « mécanismes de recherche automatique et de suppression des contenus inappropriés ». Parmi les différents outils automatisés à sa disposition, il utilise par exemple les empreintes numériques pour localiser rapidement les contenus nuisibles. Toute personne peut signaler un contenu illégal, insultant ou trompeur en cliquant sur le bouton Report (signaler). L'équipe de modération intervient le plus rapidement possible pour exclure les auteurs d'infraction et bloquer les contenus contraires aux règles de la plateforme ou à la législation en vigueur. VK permet aussi aux utilisateurs de créer des communautés et d'en devenir administrateurs et modérateurs. D'après les Conditions générales d'utilisation, il est de la responsabilité des administrateurs et des modérateurs d'une communauté de modérer et de bloquer les contenus mis en ligne sur les pages contrôlées par leur communauté. Ils doivent en particulier supprimer tout contenu contraire aux Conditions générales d'utilisation ou à la législation en vigueur. Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus discutables est probablement très faible (bien que les coûts fixes puissent être

	élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. Les utilisateurs modérateurs n'engendrent aucun coût pour VK. VK n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	En cas de violation des Conditions générales d'utilisation, y compris lors de la création et de l'administration d'une communauté, VK peut retirer/supprimer les contenus en infraction, bloquer provisoirement l'accès de l'utilisateur concerné aux services, exclure le contenu des résultats des recherches ou résilier le compte concerné.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Des comptes de l'EIIL ont été trouvés sur VK (Lokot, 2014[139]).

35. YY Live

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de YY Live disposent toutefois que les utilisateurs ne peuvent pas publier, transmettre, diffuser et conserver des contenus violents.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://zc.yy.com/license.html
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les	Non.

Conditions générales d'utilisation ou les Règles de la communauté ?	
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Aucune procédure n'est indiquée.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Pas d'informations communiquées. Des recherches ont toutefois montré que YY Live applique une surveillance et une censure par mot clé. (Knockell, 2015[140]) Pour exécuter ses Conditions générales d'utilisation, YY Live possède dans son service de sécurité des données une équipe qui assure une surveillance 24 heures/24 des contenus. Elle s'appuie sur un système qui balaie la plateforme pour détecter les contenus inappropriés et filtre automatiquement les mots clés. (Knockell, 2015[140]) Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus discutables est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. YY Live n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	En cas de violation de ses Conditions générales d'utilisation, YY Live peut limiter ou bloquer l'accès de l'utilisateur concerné à son compte, restreindre ou suspendre son accès à un ou plusieurs produits, services ou fonctions (les vidéos en direct, par exemple).
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.

8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

36. Medium

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de Medium disposent toutefois que la plateforme interdit les contenus ou actions qui menacent, encouragent ou prônent la violence, de manière directe ou indirecte, les contenus qui promeuvent la violence ou la haine sur la base de critères tels que la race, l'ethnie, l'origine nationale, la religion, le handicap, la maladie, l'âge, l'orientation sexuelle, le sexe ou l'identité sexuelle, les posts ou les comptes qui glorifient, célèbrent, minimisent ou banalisent la violence, la souffrance, les agressions ou la mort de personnes ou de groupes, ainsi que les appels à l'intolérance, l'exclusion ou la ségrégation sur la base de critères précis. L'apologie des groupes se livrant à ces comportements est également interdite.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://medium.com/policy/medium-rules-30e5502c4eb4 et https://medium.com/policy/medium-terms-of-service-9db0094a1e0f
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des	Pour les contenus signalés par les utilisateurs, Medium prend en compte des facteurs tels que l'intérêt, le contexte et la nature des informations postées, la probabilité raisonnable, l'ampleur et la

Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	gravité probables du préjudice social prévisible et la législation en vigueur. Lors de l'évaluation des contenus controversés et extrémistes (pas spécifiquement les contenus extrémistes violents) au regard des Règles, les modérateurs employés par Medium appliquent une analyse des risques qui répond au minimum aux questions suivantes : - Quelles sont les conséquences négatives prévisibles de la diffusion de l'information par Medium et de son partage sur d'autres réseaux sociaux ? - Quelle pourrait être la gravité de ses répercussions ? - Quelle est la probabilité que les conséquences négatives se produisent ? - Qui toucheront-elles probablement ? - Des informations provenant d'institutions nationales ou internationales reconnues (telles que l'ECDC, l'OMS et d'autres organes officiels) peuvent-elles nous aider à déterminer si le contenu présente un risque élevé ? (Medium, s.d.[141]) Medium fournit les exemples suivants de types de contenus présentant un risque élevé, qui sont donc davantage susceptibles d'être suspendus ou soumis à une diffusion limitée : théories du complot ayant déjà donné lieu à des incidents de harcèlement ou de violence entre les utilisateurs, théories dont on peut prévoir qu'elles peuvent encourager ou provoquer le harcèlement ou des préjudices physiques ou de réputation. (Medium, s.d.[141])
4.1 Notifications des suppressions ou des autres décisions d'exécution	Medium avertit l'utilisateur lorsqu'une enquête est menée sur un contenu de son compte, ou qu'il est désactivé, sauf s'il pense qu'il s'agit d'un compte robotisé ou opérant de mauvaise foi, ou que la notification pourrait causer un préjudice à quelqu'un, l'entretenir ou l'exacerber.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Si l'utilisateur pense que son contenu ou son compte a été limité ou désactivé par erreur, ou que Medium n'a pas eu connaissance d'un élément de contexte particulier lors de sa prise de décision, il peut déposer un recours.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler des contenus ou des comptes qui enfreignent les règles de Medium ou remplir un signalement qui décrit l'infraction présumée. Les posts et les comptes signalés sont examinés par l'équipe Confiance et sécurité (Trust & Security) de Medium, qui détermine si des règles ont été effectivement violées et prend les mesures nécessaires le cas échéant. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. Medium n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.

6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La violation des règles peut entraîner l'envoi d'avertissements, des restrictions de compte, une limitation de la diffusion des posts et des contenus, la suspension des contenus et la suspension du compte en infraction. Les contenus controversés et extrémistes (mais pas spécifiquement les contenus extrémistes violents) risquent particulièrement de faire l'objet d'une suspension ou d'une limitation de leur diffusion.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. Medium a publié en 2015 un rapport de transparence (Medium, 2015[142]) sur les demandes d'information ou de suppression de contenus émises par les autorités en 2014, mais qui ne mentionnent pas précisément les contenus terroristes et extrémistes violents.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

37. Haokan

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de Haokan interdisent toutefois d'utiliser les services de la plateforme pour fournir un soutien marqué ou des ressources à des opérations terroristes.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page http://www.haokan88.live/term_condition.html
3. Existe-t-il des dispositions précises applicables aux	Non.

contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Haokan indique qu'il a le droit (mais pas l'obligation) de bloquer ou de retirer les contenus postés sur sa plateforme à sa seule discrétion, notamment s'ils enfreignent ses Conditions générales d'utilisation.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Pas d'informations communiquées. Haokan n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Haokan indique qu'en cas de violation de ses Conditions générales d'utilisation, il est autorisé à résilier ou restreindre l'accès de l'utilisateur concerné à son compte et de supprimer les contenus en infraction sans avertissement préalable.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.

8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

38. Odnoklassniki

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation d'Odnoklassniki interdisent toutefois la propagande ou l'apologie de la haine ou du suprémacisme basée sur des critères sociaux, raciaux, nationaux ou religieux, les contenus comprenant des menaces ou incitant à la violence ou à commettre des infractions pénales et la publication d'informations à caractère extrémiste. Le terme « extrémiste » n'est pas défini.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://ok.ru/regulations
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et	Odnoklassniki indique qu'il peut avertir, notifier ou informer les utilisateurs du non-respect de ses Conditions générales d'utilisation. Ils doivent dans ce cas obligatoirement suivre les instructions qui leur sont données. Odnoklassniki explique aussi qu'il peut supprimer les contenus qui selon lui enfreignent ou peuvent enfreindre la législation en vigueur, ses Conditions générales d'utilisation, portent préjudice ou sont susceptibles de la faire, ou menacent la sécurité de ses

existe-t-il des mécanismes de recours ?	utilisateurs ou de tiers.
4.1 Notifications des suppressions	Odnoklassniki notifie les utilisateurs de leurs infractions aux Conditions générales d'utilisation à sa seule discrétion.
4.2 Mécanismes de recours en cas de suppression	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Des utilisateurs peuvent devenir modérateurs des pages personnelles d'autres utilisateurs ou créer des groupes dont ils sont ensuite modérateurs. Ils sont dans ce cas tenus de modérer les contenus publiés sur ces pages et dans ces groupes. Ils peuvent aussi devenir modérateurs de vidéos et de photos en téléchargeant l'application de modération d'Odnoklassniki (Odnoklassniki, s.d.[143]). Les utilisateurs peuvent signaler les contenus contraires aux Conditions générales d'utilisation d'Odnoklassniki. L'équipe de la plateforme examine les signalements et décide des mesures à prendre. Le coût économique marginal d'un recours à des modérateurs humains employés pour détecter les contenus répréhensibles est probablement relativement élevé. Les utilisateurs modérateurs n'engendrent aucun coût pour Odnoklassniki. Odnoklassniki indique qu'il ne censure pas de manière automatisée, et qu'il n'a pas la capacité technique de le faire, les informations publiées dans les parties accessibles au public de sa plateforme ou dans les pages personnelles des utilisateurs, ni les messages personnels. Il n'effectue pas non plus de modération en amont des informations et des contenus publiés par les utilisateurs. Odnoklassniki n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	En cas de violation de ses Conditions générales d'utilisation, Odnoklassniki a le droit de suspendre, restreindre ou supprimer l'accès de l'utilisateur concerné à sa plateforme.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les	Sans objet.

informations et données figurant dans les rapports de transparence.	
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Des contenus terroristes et extrémistes violents soutenant l'État islamiste ont été trouvés sur Odnoklassniki (Powell, 2019[144]).

39. Discord

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. La Charte d'utilisation de la communauté Discord interdit toutefois de partager des contenus qui menacent directement l'état physique ou financier d'une personne ou qui menacent de porter atteinte à une autre personne ou à une personne liée à une autre à quelque titre que ce soit. Les Conditions générales d'utilisation de Discord précisent en outre qu'il est interdit de diffamer, calomnier, tourner en ridicule, moquer, traquer, harceler, intimider ou maltraiter quiconque.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://discordapp.com/terms et https://discordapp.com/guidelines
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Discord indique que le partage de contenus qui menacent l'état physique ou financier d'une personne est totalement inacceptable et entraîne le retrait immédiat des contenus et la suppression du compte.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.

4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Les utilisateurs peuvent faire appel des décisions prises à l'encontre de leur compte.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les contenus contraires à la Charte d'utilisation de la communauté et aux Conditions générale d'utilisation. Discord indique que bien qu'il ne lise pas les messages privés des utilisateurs, il mène des enquêtes et prend des mesures immédiates si nécessaire en cas de signalement d'une violation de ses Conditions générales d'utilisation par un serveur (espace comparable à un groupe ou une communauté, créé autour d'un thème commun) (Liao, 2018[145]). Après le signalement, l'équipe Confiance et sécurité de Discord mène une enquête en examinant tous les éléments disponibles et en rassemblant le plus d'informations possible. L'enquête porte avant tout sur les messages publiés, mais elle peut être élargie si des éléments révèlent une infraction plus grande – par exemple, le serveur a pour seul but de se livrer à un comportement répréhensible ou ce comportement semble s'être déjà produit par le passé. Discord utilise des « ordinateurs intelligents », des outils automatisés et des systèmes comme PhotoDNA pour détecter le spam et les contenus d'exploitation, tels que la vengeance pornographique, l'hypertrucage et les contenus menaçant la sécurité des enfants. Il n'est toutefois pas précisé si Discord emploie ces systèmes pour détecter les contenus terroristes et extrémistes violents. Discord a reçu des signalements concernant des serveurs qui visaient principalement à diffuser des discours de haine, à harceler autrui et à faire l'apologie d'idéologies dangereuses. Il affirme prendre ces signalements au sérieux et supprimer les serveurs affichant un comportement extrémiste (pas spécifiquement extrémiste violent). Il déclare également travailler en collaboration avec les autorités répressives, des tiers (tels que des organes d'information et des universitaires) et des organisations spécialisées dans la lutte contre la haine (comme l'Anti-Defamation League ou le Southern Poverty Law Center) pour se tenir informé de tous les risques potentiels. Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus discutables est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. Discord n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions	Si Discord détecte une violation de sa Charte d'utilisation de la communauté, il peut prendre l'une des mesures suivantes à l'encontre des utilisateurs et/ou des serveurs concernés :

générales d'utilisation ou des Règles de la communauté.	- supprimer des contenus, - avertir l'utilisateur et lui expliquer l'infraction commise, - exclure provisoirement l'utilisateur pour apaiser les choses, - exclure définitivement un utilisateur et rendre difficile la possibilité de créer un nouveau compte, - supprimer un serveur, - désactiver la capacité d'un serveur à inviter de nouveaux utilisateurs.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. Discord a toutefois publié récemment un premier rapport de transparence (Discord, 2019[146]), qui comprend le nombre de signalements reçus pour des violations à la Charte d'utilisation de la communauté, qui peuvent inclure les contenus terroristes et extrémistes violents.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Le rapport couvre la période du 1er janvier au 1er avril 2019. Il indique le nombre total de signalements reçus, ainsi que le pourcentage relevant de la catégorie « comportements menaçants », qui est celle se rapprochant le plus des contenus terroristes et extrémistes violents. Il comprend aussi le pourcentage de signalements de comportements menaçants pour lesquels Discord a pris des mesures, mais ne précise pas si celles-ci ont consisté à retirer le contenu, envoyer un avertissement ou supprimer un compte. Il est important de noter que Discord reconnaît avoir reçu des signalements relatifs à la tuerie de Christchurch diffusée en direct le soir de l'attaque. Il déclare s'être employé à retirer la vidéo le plus rapidement possible des espaces où les utilisateurs pouvaient l'avoir partagée. Il a ensuite constaté une augmentation des signalements de « contenu affilié ». Il a pris de mesures pour retirer les utilisateurs qui faisaient l'apologie de l'attentat ou se faisaient passer pour le tireur, ainsi que les serveurs consacrés à une étude détaillée du manifeste du tireur ou qui soutenaient son programme, et les mèmes sur la tuerie. Discord indique le nombre approximatif de signalements sur des contenus associés à la tuerie reçus au cours des 10 jours suivants, ainsi que le nombre de comptes exclus et de serveurs supprimés pour avoir violé ses règles.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Pas d'informations communiquées.
10. Fréquence de publication des rapports de transparence.	Non définie.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Voir la section 8 ci-dessus.

40. Smule

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Directives communautaires de Smule interdisent toutefois la publication de contenus incitant à la violence.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://www.smule.com/fr/s/communityguidelines et https://www.smule.com/fr/termsofservice
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Smule indique qu'il n'effectue aucun filtrage de contenu en amont, mais se réserve le droit de retirer ou de supprimer tout contenu à sa seule discrétion, avec ou sans avertissement préalable, en particulier s'il ne respecte pas ses Directives communautaires ou Conditions générales d'utilisation. S'il trouve un « contenu répréhensible », il prend des mesures appropriées, telles que avertir l'utilisateur, suspendre ou résilier son compte, supprimer ses contenus et/ou signaler l'utilisateur aux autorités répressives, de manière directe ou indirecte.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Les notifications prennent la forme d'avertissements envoyés à la discrétion de Mule.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage	Les utilisateurs peuvent signaler les contenus contraires aux Directives communautaires et aux Conditions générales d'utilisation. Smule examine les contenus signalés par ses membres et peut les supprimer s'ils sont jugés inappropriés ou dangereux pour la communauté ou s'ils ne respectent pas de quelque manière que ce soit ses Directives communautaires ou Conditions générales d'utilisation.

d'empreintes numériques/d'adresses URL)	Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. Smule n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Si un utilisateur enfreint ses Conditions générales d'utilisation ou ses Directives communautaires, Smule peut lui envoyer un avertissement, supprimer le contenu inapproprié, résilier définitivement son compte, notifier les autorités répressives ou engager des poursuites à son encontre.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

41. KaKaoTalk

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de KaKaoTalk interdisent toutefois les contenus violents et les comportements qui encouragent les activités illégales. KaKaoTalk interdit également toute forme de discrimination qui promeut des stéréotypes fondés sur une région, le handicap, la race, l'ethnie, le sexe, l'âge, le travail et la religion.

2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://www.kakao.com/en/terms and https://www.kakao.com/policy/oppolicy?lang=en
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	KakaoTalk indique qu'en cas de violation de ses règles ou de la législation en vigueur, il peut mener des enquêtes sur les infractions, supprimer les posts concernés à titre provisoire ou définitif ou restreindre tout ou partie de ses services provisoirement ou définitivement. Le caractère provisoire ou définitif de la restriction dépend du nombre cumulé d'infractions. Cependant, toute activité interdite en vertu de la législation ou des réglementations en vigueur entraîne une restriction des services immédiate et définitive, quel que soit le nombre cumulé d'infractions.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Les utilisateurs sont avertis dans les meilleurs délais des mesures d'exécution mentionnées ci-dessus par courriel ou par d'autres outils disponibles sur l'application, sauf lorsqu'il faut protéger de manière urgente d'autres utilisateurs.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Les utilisateurs peuvent faire appel des mesures prises. Ils sont informés de la décision définitive de KakaoTalk lorsque leur demande de recours a été examinée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent créer une « chaîne de story », en devenir « maîtres » et y inviter des « directeurs ». Les maîtres et les directeurs sont les administrateurs et les modérateurs des chaînes. Ils peuvent bloquer et signaler les utilisateurs et les contenus contraires aux règles de KaKaoTalk. Les utilisateurs peuvent également signaler les contenus contraires aux règles de l'application. L'équipe de KaKaoTalk examine les signalements et prend des mesures le cas échéant. Les autorités de réglementation sud-coréennes, telles que la National Policy Agency (NPA), les Communications Commissions et la Korean Communications Standards Commission (KCSC), peuvent demander la suppression des informations antisociales, violentes et illégales. KaKaoTalk peut par ailleurs appliquer des restrictions en cas d'activités interdites en vertu de ses propres règles ou qui enfreignent la législation ou les réglementations en vigueur, sans avoir reçu de signalements de la part des utilisateurs ou des autorités de réglementation.

	Il surveille les contenus des chaînes de story, ainsi que ceux des blogs et du réseau social, à partir de mots clés en rapport avec les contenus terroristes et extrémistes violents et illégaux. KaKao TV, la plateforme de streaming, est également surveillée, de même que le contenu diffusé en direct. Si la surveillance détecte un contenu problématique, tel que du terrorisme ou de l'extrémisme violent, KaKao TV demande à l'utilisateur qui l'a mis en ligne de le modifier (c'est-à-dire de le supprimer ou de le transformer). S'il n'est pas modifié dans un délai de trois jours, il est supprimé par les modérateurs, qui excluent en outre l'utilisateur à titre provisoire ou définitif en fonction du degré de violence du contenu et du nombre cumulé d'infractions commises par l'utilisateur. Cependant, s'il est jugé que le contenu nécessite une intervention urgente, les modérateurs sont autorisés à le supprimer immédiatement. Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus discutables est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. Les utilisateurs modérateurs n'engendrent aucun coût pour KaKaoTalk. KaKaoTalk n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	En cas de violation de ses règles, KaKaoTalk peut envoyer un avertissement, supprimer les contenus en infraction, et restreindre ses services à titre provisoire ou définitif selon le nombre cumulé d'infractions. Cependant, toute activité interdite en vertu de la législation ou des réglementations en vigueur entraîne une restriction des services immédiate et définitive, quel que soit le nombre cumulé d'infractions.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. KaKaoTalk publie des rapports de transparence (Daum Kakao, s.d.[147]) qui comprennent les demandes émises par les autorités sud-coréennes pour accéder aux informations des utilisateurs et supprimer des contenus, mais qui ne mentionnent pas précisément les contenus terroristes et extrémistes violents.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des	Information inconnue.

contenus terroristes et extrémistes violents ?	

42. DeviantArt

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de DeviantArt prévoient toutefois que les commentaires agressifs ou insultants sans raison sont interdits (Prohibited Commentaries). Les utilisateurs ne doivent pas utiliser DeviantArt à des fins illégales ou pour mettre en ligne, publier ou diffuser des contenus illégaux, menaçants, nuisibles ou répréhensibles de quelque manière que ce soit.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page https://about.deviantart.com/policy/service/, https://about.deviantart.com/policy/etiquette/ et https://about.deviantart.com/policy/submission/
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Lorsqu'un contenu interdit (appelé « écart ») est signalé, son auteur peut recevoir une notification anonyme lui demandant si le contenu est par exemple un « contenu adulte », selon le motif du signalement. L'auteur a ainsi la possibilité d'intervenir et de remédier au problème. S'il n'intervient pas et que le contenu ne fait l'objet d'aucun nouveau signalement, l'équipe de DeviantArt peut décider qu'il n'est pas nécessaire de le supprimer ou de lui apposer une mention et invalider le signalement. En revanche, si le nombre de signalements augmente, le personnel traite le contenu en priorité et prend les mesures nécessaires plus rapidement, en lui ajoutant une mention ou en le supprimant, ou en invalidant le signalement. Il convient de souligner que même si une notification est envoyée à l'utilisateur concerné, les signalements sont systématiquement transmis au personnel de DeviantArt pour approbation. L'envoi de la notification permet simplement de laisser à l'utilisateur la possibilité de corriger une éventuelle erreur commise de bonne foi (Kitsune, 2017[148]). L'utilisation des outils de communication fournis par DeviantArt à des fins délibérément agressives ou inappropriées peut entraîner l'application de mesures disciplinaires (DeviantArt, s.d.[149]). Les fils de discussion des forums qui contiennent des propos déplacés, portent sur des sujets inappropriés ou comprennent un nombre indésirable d'infractions aux règles de DeviantArt sont bloqués et fermés aux commentaires. Un utilisateur inscrit sur DeviantArt peut être administrateur ou membre d'un « groupe », c'est-à-dire un ensemble d'applications

	et de pages d'utilisateurs créé pour rassembler des contenus, des discussions et des membres autour d'intérêts communs. Les administrateurs peuvent définir leurs propres règles et les droits accordés aux utilisateurs/membres de leur groupe. D'une manière générale, DeviantArt n'intervient pas dans les groupes, sauf en cas de violation manifeste de ses règles. Il peut alors supprimer le groupe concerné et les droits qui lui sont associés. Les utilisateurs qui ont fait preuve d'un comportement inacceptable en ne respectant pas les règles de DeviantArt ou en se livrant à des activités inappropriées ou préjudiciables à la communauté peuvent faire l'objet d'une suspension provisoire de leur compte. (DeviantArt, s.d.[150]) Si un compte est suspendu, le message « compte suspendu » s'affiche pendant toute la durée de la suspension à la place de la page de profil habituelle de l'utilisateur. La durée des suspensions administratives peut varier. Elles sont le plus souvent imposées pour 24 heures, une semaine, deux semaines ou 30 jours. Pendant ce temps, l'utilisateur ne peut plus envoyer de posts, utiliser la plupart des fonctionnalités de la plateforme ou interagir avec la communauté en général. L'utilisateur reçoit une notification de la mesure prise, qui peut être assortie d'un message privé ou du motif de la mesure, et un minuteur est ajouté sur la page de son profil. Si l'utilisateur se voit ensuite imposer des sanctions disciplinaires, les éventuelles suspensions dont il a fait l'objet par le passé sont prises en compte. Son compte peut être suspendu pour une durée plus longue, ou, en cas d'infractions répétées, une nouvelle suspension peut entraîner la résiliation du compte (DeviantArt, s.d.[151]).
4.1 Notifications des suppressions ou des autres décisions d'exécution	Si un contenu est supprimé par le personnel de DeviantArt, son auteur reçoit une notification. Les suspensions de compte sont également notifiées.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Si l'auteur pense que son contenu est autorisé et que le personnel a commis une erreur, il peut contester la décision en expliquant ses raisons. Le personnel examinera alors une nouvelle fois le contenu. DeviantArt autorise généralement ses utilisateurs à déposer un recours et à demander des précisions sur des suppressions de contenu, des avertissements pour infractions, des suspensions et résiliations de compte ou toute autre mesure administrative. Les recours, demandes de renseignements et questions sont examinés et traités par le personnel de DeviantArt.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les administrateurs de groupe sont modérateurs au sein de leur groupe. Les utilisateurs peuvent également signaler les contenus contraires aux règles de l'application. Lorsque le personnel de DeviantArt est informé d'une infraction, il examine le signalement et prend les mesures nécessaires le cas échéant. DeviantArt indique ne pas avoir la possibilité de contrôler les contenus que les utilisateurs peuvent mettre en ligne, publier ou

	diffuser sur sa plateforme et ne pas être soumis à l'obligation de surveiller ces contenus pour quelque raison que ce soit. Le coût économique marginal d'un recours à des modérateurs humains employés pour détecter les contenus répréhensibles est probablement relativement élevé. Les utilisateurs modérateurs n'engendrent aucun coût pour DeviantArt. DeviantArt n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La violation des Conditions générales d'utilisation peut entraîner l'envoi d'un avertissement, la suppression du contenu, la suspension du compte ou la résiliation de l'adhésion de l'utilisateur, à la discrétion de DeviantArt.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Des groupes néo-nazis ont utilisé DeviantArt pour mettre en ligne de la propagande et recruter des membres (Hayden, 2019[152]).

43. Meetup

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Cependant, au titre de la Politique d'utilisation et de contenu de Meetup, il est interdit de publier des contenus violents explicites ; les comportements incitant à la violence envers des individus ou des groupes d'individus, sur la simple base de leurs origines ou de leurs croyances sont interdits ; et il est interdit d'utiliser Meetup pour promouvoir, faciliter ou organiser des activités à caractère violent, criminel ou non consenties, qui mettraient en péril autrui, que ce soit physiquement, psychologiquement ou émotionnellement. Par ailleurs les « groupes » (espaces créés autour d'intérêts ou d'activités précis) ne doivent pas contenir ou promouvoir des événements qui organisent, encouragent, proposent ou diffusent des services ou recrutent pour des organisations terroristes, comprendre des contenus ou promouvoir des événements susceptibles de menacer la sécurité du public ou des personnes, tels que des encouragements, des incitations ou des déclarations intentionnelles ou des menaces de commettre un acte de violence à l'encontre d'un groupe, d'une personne ou d'un lieu, faire l'apologie des armes et de la fabrication d'explosifs, et inciter à la violence en réaction à des événements privés ou publics.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://help.meetup.com/hc/fr-fr/articles/360002897532-Politique-d-utilisation-et-de-contenu, https://help.meetup.com/hc/en-us/articles/360004285732-Meetup-social-media-community-standards, https://help.meetup.com/hc/fr-fr/articles/360002897712-Politique-concernant-les-groupes-Meetup-normes-et-standards- et https://help.meetup.com/hc/fr-fr/articles/360027447252
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Meetup prévoit qu'une violation de ses règles et Conditions générales d'utilisation peut entraîner la modification, la suspension ou la résiliation du compte ou de l'accès à la plateforme et le cas échéant, il informe l'utilisateur concerné des motifs de la modification, de la suspension ou de la résiliation.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Les utilisateurs sont informés des décisions d'exécution.

4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Si un utilisateur pense que son compte a été modifié, suspendu ou résilié par erreur, il peut faire appel de la décision.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les administrateurs des groupes modèrent les contenus de leurs groupes et peuvent modifier, suspendre ou résilier l'accès des utilisateurs à ces groupes. Les utilisateurs peuvent également signaler les contenus contraires aux règles de la plateforme. L'équipe Confiance et sécurité de Meetup examine les signalements reçus et prend des mesures le cas échéant. Le coût économique marginal d'un recours à des modérateurs humains employés pour détecter les contenus répréhensibles est probablement relativement élevé. Les utilisateurs modérateurs n'engendrent aucun coût pour Meetup. Meetup indique qu'il n'examine généralement pas les contenus avant qu'ils soient publiés (Meetup, 2019[153]). Meetup n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La violation des règles de Meetup peut entraîner la suppression du contenu, la modification, la suspension ou la résiliation du compte de l'utilisateur concerné.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. Meetup publie des rapports de transparence (Meetup, 2017[154]) qui comprennent les demandes d'accès aux informations des utilisateurs émanant des autorités et les demandes de suppression de contenu en raison d'atteinte aux droits de la propriété intellectuelle, mais qui ne mentionnent pas les contenus terroristes et extrémistes violents.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

44. 4chan

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de 4chan interdisent toutefois les contenus qui enfreignent la législation locale ou des États-Unis.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles à la page http://www.4chan.org/rules#global4
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Selon 4chan, les fils de discussion arrivent à expiration ou sont supprimés par le logiciel de la plateforme relativement rapidement. La plupart des tableaux (thématiques de fils de discussion) étant limités à 10 pages, les contenus ne peuvent généralement être vus que pendant quelques heures ou quelques jours avant d'être retirés. Les posts sont la plupart du temps retirés automatiquement, mais ils peuvent parfois l'être par un modérateur ou « gardien ». Les modérateurs sont des personnes choisies pour assurer la maintenance générale de la plateforme. Ils peuvent supprimer des posts publiés sur toute la plateforme, exclure des utilisateurs, fermer des fils de discussions et prendre des mesures associées. Les « gardiens » constituent une catégorie située entre l'utilisateur et le modérateur. Ils ont accès au système de signalement de la plateforme et peuvent supprimer des posts sur les tableaux qui leur sont confiés et envoyer des demandes d'exclusion. Ils sont sélectionnés sur candidature, à l'issue d'un processus de tests et d'intégration. L'entrée dans l'équipe de modération se fait uniquement sur invitation. Le programme des gardiens accueille de nouveaux participants de temps à autre. Il n'est conservé aucun enregistrement public des contenus supprimés et les fils de discussion étant retirés régulièrement, il n'existe aucun moyen de savoir quels contenus ont été supprimés par l'équipe de modération. Autrement dit, l'utilisateur ne peut absolument pas connaître précisément sur quels contenus porte la modération ni à quel moment elle est effectuée. Les modérateurs se réservent le droit de bloquer ou d'interdire l'accès d'un utilisateur ou de supprimer des contenus pour quelque raison que ce soit sans avertissement préalable. Les utilisateurs ne peuvent plus provisoirement publier de posts

	lorsqu'une demande d'exclusion de leur adresse IP est en attente de traitement. Le blocage dure 15 minutes à partir du moment où un gardien envoie la demande d'exclusion et est levé immédiatement si un modérateur rejette la demande. Si en revanche si celle-ci est approuvée, l'exclusion est appliquée.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	S'ils pensent qu'une erreur a été commise, les utilisateurs peuvent déposer un recours en contactant les modérateurs.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	4chan indique qu'il encourage le signalement des posts (4chan, s.d.[155]). Ils sont ensuite examinés par les modérateurs, qui prennent des mesures le cas échéant. Le coût économique marginal d'un recours à des modérateurs humains employés pour détecter les contenus répréhensibles est probablement relativement élevé. Les utilisateurs modérateurs n'engendrent aucun coût pour 4chan. 4chan n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La violation des règles peut entraîner la suppression du post, une exclusion provisoire ou dans certains cas, définitive.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.

11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. La propagande néo-nazi est par exemple fréquente sur 4chan (Arthur, 2019[156]).

45. MySpace

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de MySpace interdisent toutefois l'« extrémisme malveillant », qui comprend les contenus qui prônent la violence ou la haine à l'encontre de personnes ou de groupes sur la base de la race, de l'origine ethnique, de la religion, du handicap, du sexe, de l'âge, du statut d'ancien combattant ou de l'orientation sexuelle. Il est strictement interdit d'inciter à la violence ou à porter préjudice à autrui.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://myspace.com/pages/terms#3 et https://help.myspace.com/hc/en-us/articles/202579130-Myspace-Guidelines
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	MySpace indique qu'il se réserve le droit de mener des enquêtes et de prendre des mesures (pouvant inclure des poursuites judiciaires) à l'encontre des personnes qui, selon son entière discrétion, enfreignent ses Conditions générales d'utilisation, et notamment mais pas exclusivement, supprimer de la plateforme le contenu en infraction, résilier l'adhésion de la personne concernée et/ou signaler le contenu ou les activités en infraction aux autorités répressives. MySpace peut vouloir obtenir des informations auprès de l'utilisateur soupçonné d'avoir enfreint ses Conditions générales d'utilisation et auprès d'autres utilisateurs et coopère pleinement avec les autorités répressives ou applique les décisions judiciaires lui demandant de communiquer l'identité d'une personne envoyant des courriels, publiant ou mettant à disposition tout type de contenu présumé contraire à ses Conditions générales d'utilisation.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.

5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les contenus et les profils qu'ils pensent contraires aux règles de MySpace. Une équipe de MySpace examine les signalements et peut prendre contact avec l'utilisateur qui les a envoyés pour obtenir des informations complémentaires avant de prendre une décision. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. MySpace n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La violation des Conditions générales d'utilisation peut entraîner un blocage ou une suppression du contenu et MySpace peut décider de supprimer le profil de l'utilisateur concerné s'il estime que le contenu a été publié dans le but d'inciter à la violence ou de nuire à autrui.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. MySpace a été considéré un temps comme un espace de recrutement des terroristes aux États-Unis (Farrell, 2006[157]).

46. Google Drive

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation	Il n'existe pas de définition particulière des contenus terroristes et extrémistes violents. Cependant, le Règlement du programme concernant l'utilisation abusive de Google (Google, s.d.[158]), qui s'applique aussi à Google Drive, comprend des dispositions particulières sur la violence, les incitations à la haine et les activités à caractère terroriste. *Violence* : il est interdit de menacer une personne de violences graves ou de mort, ou d'obtenir l'appui d'autres personnes dans le but de lui porter physiquement atteinte. En cas de menace grave et imminente de blessure

ou les Règles de la communauté ?	physique ou de mort, Google est susceptible de prendre les mesures nécessaires concernant le contenu. La violence gratuite ou les contenus à caractère sanglant visant à choquer ou à susciter un intérêt malsain ne sont pas autorisés. S'ils publient des contenus explicites dans un contexte documentaire, scientifique, artistique ou en lien avec l'actualité, les utilisateurs doivent donner suffisamment d'informations pour que les autres personnes comprennent ce dont il s'agit. Dans certains cas, les contenus publiés peuvent être si violents ou choquants qu'aucun contexte ne pourra justifier leur présence sur les plateformes de Google. Les utilisateurs ne doivent pas encourager les autres personnes à commettre des actes de violence spécifiques. *Incitation à la haine* : l'incitation à la haine est interdite. L'incitation à la haine désigne tout contenu qui incite à la violence ou la justifie, ou dont l'objectif principal est d'inciter à la haine envers une personne ou un groupe en raison de son origine ethnique, de sa religion, de son handicap, de son âge, de sa nationalité, de son statut d'ancien combattant, de son sexe, de son orientation ou identité sexuelle ou de toute autre caractéristique associée à une discrimination ou une marginalisation systématique. *Activités à caractère terroriste* : les organisations terroristes ne sont pas autorisées à utiliser Google Drive à quelque fin que ce soit, y compris pour le recrutement. Les contenus à caractère terroriste, tels que la promotion d'actes terroristes, l'incitation à la violence ou l'apologie d'attentats terroristes, sont aussi formellement interdits. Le terme « organisations terroristes » n'est pas défini. Si les utilisateurs publient des contenus liés au terrorisme dans un objectif pédagogique, documentaire, scientifique ou artistique, ils doivent donner suffisamment d'informations pour que les autres personnes comprennent le contexte.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://www.google.com/drive/terms-of-service/ et https://support.google.com/docs/answer/148505?visit_id=637064013896463652-1393240150&rd=1
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Non.
4. Politiques et procédures d'application et d'exécution des	Si des fichiers font l'objet d'un signalement pour infraction, un drapeau peut s'afficher à côté de leur nom et leur propriétaire ne peut plus les partager. Les fichiers ne sont plus accessibles publiquement, même pour les utilisateurs qui disposent du lien. S'ils pensent que leurs fichiers n'enfreignent pas les Conditions d'utilisation ou le Règlement du programme Google Drive, ils

Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	peuvent demander qu'ils soient examinés (Google, s.d.[159]). Si un utilisateur enfreint de façon significative ou répétée les Conditions ou le Règlement du programme, Google est susceptible d'interrompre ou de désactiver définitivement son accès à Google Drive. Dans ce cas, l'utilisateur en est informé à l'avance. Google peut toutefois être amené à suspendre ou à désactiver l'accès d'un utilisateur à Google Drive sans préavis s'il utilise ce service d'une façon susceptible de mettre en jeu la responsabilité légale de Google ou d'empêcher d'autres personnes d'utiliser Google Drive.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les contenus contraires aux Conditions générales d'utilisation et au règlement de Google Drive. Les signalements sont examinés par le personnel de Google. Google indique que l'envoi d'un signalement ne garantit pas que le fichier concerné sera supprimé ni qu'il prendra des mesures en conséquence. Les contenus qui déplaisent à un utilisateur ou qui lui semblent inappropriés n'enfreignent en effet pas forcément les Conditions générales d'utilisation ou le Règlement du programme. Google précise également qu'il est susceptible d'examiner les contenus et le comportement des utilisateurs sur Google Drive pour vérifier qu'ils respectent les Conditions générales d'utilisation et le Règlement du programme (Google, 2019[160]). La société a indiqué que les fichiers stockés sur Google Drive sont analysés par un algorithme qui détecte les infractions à son Règlement et bloque automatiquement les fichiers réputés ne pas le respecter. Ce système ne comporte pas d'examen humain (Titcomb, 2017[161]). Le coût économique marginal d'une utilisation des outils automatisés pour détecter les contenus discutables est probablement très faible (bien que les coûts fixes puissent être élevés), alors que celui d'un recours à des modérateurs humains pour la même fonction est probablement relativement élevé. Google Drive n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas	En cas de mise en ligne d'un contenu inapproprié contraire aux Conditions générales d'utilisation ou au Règlement, Google peut prendre les mesures suivantes :

de violation des Conditions générales d'utilisation ou des Règles de la communauté.	- supprimer le fichier du compte, - appliquer des restrictions au partage du fichier, - limiter l'accès au fichier, - désactiver l'accès à un ou plusieurs produits Google, - supprimer le compte Google. (Google, s.d.[162])
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. Google publie des rapports de transparence (Google, s.d.[85]) sur l'ensemble de ses produits et services, dont Google Drive. Ils comportent une partie sur les demandes de suppression de contenu émanant des autorités en raison d'infractions aux législations locales ou aux Conditions générales d'utilisation ou Règlement de Google, mais ne mentionnent pas précisément les contenus terroristes et extrémistes violents.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Des contenus de l'EIIL ont été trouvés sur Google Drive (Katz, 2018[163]).

47. Dropbox

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. La Politique d'utilisation acceptable de Dropbox dispose toutefois que les utilisateurs ne doivent pas publier ou partager du contenu contenant des actes de violence extrême, des actes terroristes, notamment de propagande terroriste.

2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://www.dropbox.com/terms et https://www.dropbox.com/terms#acceptable_use
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Sans objet.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Dropbox indique qu'il suspendra ou résiliera l'accès d'un utilisateur à ses services s'il ne respecte pas les Conditions générales d'utilisation ou utilise les services d'une manière susceptible de causer un risque réel de dommages ou de pertes pour Dropbox ou les autres utilisateurs. Il prévient l'utilisateur à l'avance dans un délai raisonnable à l'adresse électronique associée à son compte pour lui donner la possibilité d'exporter ses données. Si, après réception de cette notification, l'utilisateur ne prend pas les mesures demandées, Dropbox suspendra ou résiliera son accès aux services. Dropbox n'envoie pas de notification à l'avance si l'utilisateur commet une violation substantielle des Conditions générales d'utilisation, une violation qui risquerait d'engager sa responsabilité ou de compromettre sa capacité à fournir ses services aux autres utilisateurs ou si la loi le lui interdit.
4.1 Notifications des suppressions ou des autres décisions d'exécution	Aucune notification n'est indiquée.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Aucune procédure de recours n'est indiquée.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Les utilisateurs peuvent signaler les contenus contraires aux Conditions générales d'utilisation et aux règles de Dropbox. L'équipe de Dropbox examine les signalements, mène une enquête sur l'infraction présumée et prend des mesures le cas échéant. La société a déjà indiqué que son personnel avait besoin en de rares occasions d'accéder au contenu des fichiers des utilisateurs, en particulier pour faire appliquer ses Conditions générales d'utilisation et ses règles (Dropbox, s.d.[164]). Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. Dropbox est membre du GIFCT.
6. Sanctions ou conséquences en cas de violation des	La violation des Conditions générales d'utilisation ou des règles de Dropbox peut entraîner la suspension ou la résiliation du compte de

Conditions générales d'utilisation ou des Règles de la communauté.	l'utilisateur concerné.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. Dropbox publie des rapports de transparence (Dropbox, s.d.[165]) qui comportent une partie sur les demandes de suppression de contenu émanant des autorités en raison d'infractions aux législations locales ou aux Conditions générales d'utilisation ou règles de Dropbox, mais qui ne mentionnent pas précisément les contenus terroristes et extrémistes violents.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Des contenus de l'EIIL ont été trouvés sur Dropbox (Bennett, 2019[126]).

48. Microsoft OneDrive

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Le Contrat de services de Microsoft, qui régit OneDrive, interdit toute activité nuisible à d'autres personnes, telle que la publication de contenu terroriste ou extrémiste violent, des propos haineux ou des appels à la violence contre des tiers. Microsoft indique que, dans le cadre de ses services, un contenu terroriste désigne un contenu posté par, ou visant à soutenir, une organisation figurant sur la liste consolidée des sanctions du Conseil de sécurité des Nations Unies (Conseil de Sécurité des Nations Unies, s.d.[110]), qui représente explicitement la violence, encourage les actes violents, cautionne une organisation terroriste ou ses actes et incite à rejoindre ces groupes. La liste des sanctions des Nations Unies répertorie les groupes considérés par le Conseil de sécurité des Nations Unie comme des organisations terroristes (Microsoft, 2016[108]).
2. Manière dont les Conditions générales d'utilisation ou les	Disponibles aux pages https://www.microsoft.com/fr-fr/servicesagreement/

Règles de la communauté sont communiquées.	
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Sans objet.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	Microsoft indique qu'il se réserve le droit de supprimer ou de bloquer le contenu d'un utilisateur sur OneDrive à quelque moment que ce soit s'il est porté à sa connaissance qu'il est susceptible d'enfreindre la législation applicable ou son Contrat de service. Lors des enquêtes relatives aux infractions présumées aux Contrats de services, Microsoft se réserve le droit de consulter le contenu afin de résoudre le problème. La société précise toutefois qu'elle ne surveille pas OneDrive. Elle suit une procédure de notification et de retrait pour la suppression des contenus interdits, dont les contenus terroristes. Selon cette procédure, Microsoft reçoit une (de la part des autorités ou d'un utilisateur, par exemple), puis elle retire le contenu concerné. Si la présence de contenus terroristes sur les services aux consommateurs qu'elle héberge, dont OneDrive, est portée à sa connaissance par ses outils de signalement en ligne, elle les supprime (Microsoft, 2016[108]). Ainsi que le prévoit l'Accord de services de Microsoft, « si vous enfreignez les présentes Conditions, nous pouvons (...) cesser de vous fournir les Services ou fermer votre compte Microsoft, ou bloquer l'envoi d'une communication (ex. un e-mail, un partage de fichiers ou un message instantané) vers ou depuis les Services. Nous nous réservons également le droit de supprimer ou de bloquer votre contenu des Services à tout moment si nous pensons qu'il pourrait enfreindre la réglementation applicable ou les présentes Conditions. Lors des enquêtes relatives aux infractions suspectées des présentes Conditions, Microsoft se réserve le droit de consulter votre contenu afin de résoudre le problème. »
4.1 Notifications des suppressions ou des autres décisions d'exécution	Les notifications sont envoyées à la discrétion de Microsoft. Selon le Contrat de services de Microsoft, « Si une information doit vous être communiquée concernant un service que vous utilisez, nous vous enverrons les notifications de service (...). Si vous nous avez donné votre adresse e-mail ou votre numéro de téléphone dans le cadre de votre compte Microsoft, vous êtes susceptible de recevoir des notifications de service par e-mail ou SMS, y compris pour vérifier votre identité avant d'enregistrer votre numéro de téléphone mobile et de vérifier vos achats. Vous êtes susceptible de recevoir des notifications de service par d'autres moyens (par exemple, par des messages intégrés au produit). »

4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Le formulaire pour faire appel de la suspension d'un compte Microsoft est disponible à la page https://www.microsoft.com/en-us/concern/AccountReinstatement
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	Microsoft déploie toute une gamme d'outils de filtrage, de solutions d'intelligence artificielle, de partenariats extérieurs et d'opérations de modération humaine pour détecter les contenus terroristes et extrémistes violents et mener des enquêtes à ce sujet. Les utilisateurs peuvent signaler les contenus contraires aux politiques de Microsoft. Des modérateurs examinent les signalements pour déterminer les mesures à prendre. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. Microsoft est un membre fondateur du GIFCT et participe au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Si un utilisateur publie un contenu interdit ou qui enfreint de manière substantielle le Contrat de service, Microsoft est susceptible de prendre des mesures à son encontre, telles qu'interrompre l'accès à OneDrive, fermer immédiatement le compte Microsoft de l'utilisateur ou bloquer l'envoi des communications (courriel, partage de fichiers ou messagerie instantanée) vers ou depuis OneDrive. Microsoft peut également bloquer ou supprimer le contenu. Voir également la section 4 ci-dessus et l'article de blog publié en 2016 ci-après. « **Application de la procédure de notification et retrait** : nous continuons à appliquer la procédure de notification et retrait pour supprimer les contenus interdits, dont les contenus terroristes. Si la présence de contenus terroristes sur les services aux consommateurs que nous hébergeons est portée à notre connaissance par nos outils de signalement en ligne, nous les supprimons. Tous les signalements de contenu terroriste – effectués par les autorités, des citoyens ou d'autres groupes – publié sur un service Microsoft doivent nous être envoyés par l'intermédiaire de ce formulaire. » (https://blogs.microsoft.com/on-the-issues/2016/05/20/microsofts-approach-terrorist-content-online/)
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. Microsoft publie des rapports de transparence (Microsoft, 2019[111]) qui comportent une partie sur les demandes de suppression de contenu émanant des autorités qui signalent des violations de ses Conditions générales d'utilisation ou de la législation locale, mais qui ne mentionnent pas précisément les contenus terroristes et extrémistes violents.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Sans objet.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les	Sans objet.

informations et données figurant dans les rapports de transparence.	
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Des vidéos de l'EIIL ont été hébergées sur OneDrive (Counter Extremism Project, 2018[166]).

49. WordPress.com

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et violents, mais les Conditions générales d'utilisation de WordPress.com interdisent les sites Web des groupes terroristes reconnus comme tels par les autorités américaines. Le Bureau du contrôle des actifs étranger (Office of Foreign Assets Control, OFAC) du Trésor américain tient à jour la liste des « ressortissants spécialement désignés » (US Treasury, 2020[167]) avec lesquels WordPress.com n'a pas le droit de faire des affaires. WordPress.com interdit aux personnes, groupes ou entités figurant sur cette liste d'utiliser ses services (WordPress, s.d.[168]). Les incitations explicites à la violence sont également interdites. Cela comprend la publication de contenus qui profèrent des menaces, encouragent ou incitent à la violence ou à causer des préjudices physiques ou la mort, menacent des personnes ou des groupes ciblés ou de commettre des actes de violence aveugles.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://wordpress.com/fr/tos/ et https://en.support.wordpress.com/user-guidelines/
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Sans objet.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et	WordPress.com a travaillé en collaboration avec des spécialistes de l'extrémisme en ligne et les autorités répressives pour élaborer des mesures destinées à lutter contre la propagande terroriste et extrémiste (pas spécifiquement extrémiste violente). La société peut suspendre les sites Web qui incitent à la violence ou qui sont liés à des groupes terroristes officiellement interdits (au titre de la liste de l'OFAC du Trésor américain), quel que soit leur contenu. Elle applique également d'autres mesures. Elle peut par exemple signaler un contenu ou retirer un site du Lecteur WordPress.com

existe-t-il des mécanismes de recours ?	pour que son contenu soit plus difficile à trouver. Un site qui fait l'objet d'un signalement est automatiquement retiré de tous les programmes publicitaires exécutés par WordPress.com. D'après la société, les signalements effectués par les unités de référence internet (IRU) publiques représentent un moyen important d'attirer son attention sur des sites extrémistes (mais pas spécifiquement extrémistes violents, rappelons-le). Ces organismes possèdent en effet une expertise de la propagande en ligne dont ne peuvent pas se doter les sociétés de technologie privées. Ils s'emploient à détecter les sites utilisés par des terroristes connus pour diffuser de la propagande ou organiser des actes de violence. Ils signalent les sites terroristes à WordPress.com par le biais d'une adresse électronique spéciale qui permet à la société de repérer plus facilement les signalements provenant d'une source fiable. WordPress.com ne retire pas automatiquement les sites Web de sa plateforme. Un membre de son équipe Risque et sécurité examine tous les signalements reçus pour déterminer si les contenus concernés enfreignent les règles de la plateforme. L'une des raisons d'examiner tous les signalements est d'éviter de supprimer des contenus postés sur des sites légitimes (organes d'informations, sites universitaires) qui parlent du terrorisme ou de groupes terroristes. WordPress.com héberge les sites d'un certain nombre de très grands organes d'informations, de blogueurs d'actualité, d'universitaires et de chercheurs, qui publient tous des contenus légitimes sur le terrorisme. Dans un autre contexte, certains de ces contenus pourraient toutefois être considérés comme de la propagande terroriste, auquel cas ils seraient retirés de la plateforme au titre des règles qu'elle applique. WordPress.com indique que le contexte est très important et qu'elle ne peut confier à un robot les décisions susceptibles d'affecter un contenu légitime. Dans la mesure où le volume des signalements qu'elle reçoit n'est pas très important comparé à d'autres plateformes, elle peut effectuer plus d'examens humains que d'examens automatisés (Clicky, 2017[169]).
4.1 Notifications des suppressions ou des autres décisions d'exécution	WordPress.com explique que, en fonction du scénario, elle enverra un courriel ou ajoutera une notification d'avertissement au tableau de bord de l'utilisateur qui a enfreint ses règles. La notification contiendra un lien que l'utilisateur pourra utiliser pour contacter la plateforme au sujet du problème. Ces « scénarios » ne sont toutefois pas définis (WordPress.com, s.d.[170]).
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Les utilisateurs peuvent faire appel des mesures d'exécution appliquées par WordPress.com s'ils estiment qu'elles ont été prises par erreur. Une personne réelle examinera leur demande et leur communiquera la décision définitive dès que possible.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par les utilisateurs, évaluateurs humains	WordPress.com n'effectue pas de filtrage en amont des contenus postés par les utilisateurs. Les utilisateurs peuvent signaler les contenus ou les sites qu'ils estiment contraires aux règles de WordPress.com. En outre, comme cela a été indiqué plus haut, les unités de référence

(personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	internet (IRU) signalent les sites terroristes et extrémistes à la plateforme, qui les examine et prend des mesures le cas échéant. Le coût économique marginal d'un recours à des modérateurs humains pour détecter les contenus répréhensibles est probablement relativement élevé. WordPress.com n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	Si WordPress.com détecte un site ou un contenu contraire à ses règles, elle supprime le contenu, désactive certaines fonctionnalités du compte et/ou suspend la totalité du site.
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Oui. Automattic, la société-mère de WordPress.com, publie des rapports de transparence qui comportent une partie sur les signalements effectués par les unités de référence internet (IRU) pour des contenus extrémistes (pas spécifiquement extrémistes violents) (Automattic, s.d.[171]). Le dernier rapport comprend les données du 1er janvier au 30 juin 2019.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	- Nombre de notifications de contenu extrémiste (pas spécifiquement extrémiste violent) des unités de référence internet (IRU) - Nombre de notifications qui ont entraîné un retrait du contenu/du site - Pourcentage des notifications qui ont entraîné un retrait du contenu/du site Les chiffres sont répartis par mois (de janvier à juin) et pays de signalement.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Pas d'informations communiquées.
10. Fréquence de publication des rapports de transparence.	Semestrielle. Automattic a publié des rapports de transparence pour les périodes suivantes : - 2017 : 1er juillet – 31 décembre - 2018 : 1er janvier – 30 juin - 2018 : 1er juillet – 31 décembre - 2019 : 1er janvier – 30 juin
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Oui. Voir la section 7 ci-dessus.

50. Wikipedia

1. Comment les contenus terroristes et extrémistes violents sont-ils définis dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Il n'existe pas de définition des contenus terroristes et extrémistes violents. Les Conditions générales d'utilisation de la Fondation Wikimedia, qui régissent Wikipedia, interdisent toutefois notamment le harcèlement, les menaces, les propos outrageants et le vandalisme. Il est également interdit d'utiliser les services de Wikimedia d'une façon incompatible avec la loi applicable.
2. Manière dont les Conditions générales d'utilisation ou les Règles de la communauté sont communiquées.	Disponibles aux pages https://foundation.wikimedia.org/wiki/Terms_of_Use/fr et https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:R%C3%A8gles_et_recommandations
3. Existe-t-il des dispositions précises applicables aux contenus diffusés en direct dans les Conditions générales d'utilisation ou les Règles de la communauté ?	Sans objet.
4. Politiques et procédures d'application et d'exécution des Conditions générales d'utilisation ou des Règles de la communauté (suppression des contenus). En particulier : les suppressions ou les autres décisions d'exécution font-elles l'objet de notifications et existe-t-il des mécanismes de recours ?	La communauté Wikipedia occupe un rôle de premier plan dans la définition et l'exécution des règles. Elle se compose des fonctions suivantes : - *Contributeurs* : bénévoles qui écrivent et modifient les pages de Wikipedia. - *Stewards* : contributeurs bénévoles chargés de la mise en œuvre technique des droits des utilisateurs et habilités à utiliser la fonction CheckUser (Wikipedia, 2019[172]) et à masquer les contenus (Wikipedia, 2019[172]). - *Bureaucrates* : contributeurs volontaires habilités à accorder ou retirer à d'autres utilisateurs le statut d'administrateur ou de bureaucrate, et attribuer ou supprimer le statut de bot à un compte. - *Administrateurs* : contributeurs auxquels est confié l'accès à des fonctionnalités techniques particulières (outils). Ils peuvent par exemple protéger et supprimer des pages ou bloquer des contributeurs (Wikipedia, 2020[173]). Les principales règles de Wikipedia relatives aux contenus sont les suivantes : 1. Neutralité de point de vue : tous les articles de Wikipedia et les contenus encyclopédiques doivent être rédigés avec neutralité, en présentant les principaux points de vue de manière impartiale et proportionnelle, sans parti pris.

	2. Vérifiabilité : les personnes qui lisent et modifient l'encyclopédie peuvent vérifier que les informations proviennent d'une source fiable. 3. Pas de publication de travaux inédits : Wikipedia ne publie pas de recherches qui n'ont encore jamais été publiées. Tous les contenus publiés sur Wikipedia doivent être associés à des sources fiables et publiées (Wikipedia, 2019[174]). 4. Les contenus sont supprimés par les administrateurs s'ils estiment qu'ils enfreignent les règles de contenu ou les autres règles de Wikipedia ou la législation des États-Unis (Wikipedia, 2020[175]). 5. 6. La suppression repose sur les processus appliqués lors de la mise en œuvre et de l'enregistrement des décisions de la communauté concernant la suppression de pages et de médias (Wikipedia, 2020[176]). Une discussion doit en principe avoir lieu au préalable pour constituer un consensus favorable à la suppression. Les administrateurs sont généralement chargés de clore ces discussions, mais des contributeurs non-administrateurs en règle peuvent le faire sous certaines conditions. Des contributeurs peuvent toutefois demander la suppression d'une page s'ils pensent que celle-ci sera acceptée sans controverse. Dans certains cas, il est possible de supprimer rapidement une page si elle remplit les critères définis par la communauté, qui incluent notamment les pages créées aux seules fins de dénigrer, menacer, intimider ou harceler leur sujet ou une autre entité (Wikipedia, 2020[177]). La Fondation Wikimedia déclare intervenir rarement dans les décisions de la communauté concernant les règles et leur exécution. Cependant, si la communauté demande une intervention ou la prise en charge d'un utilisateur particulièrement problématique parce qu'il crée des troubles importants ou se livre à un comportement dangereux, elle peut mener une enquête sur l'utilisation des services par l'utilisateur (a) pour déterminer si une infraction aux règles ou à la législation a eu lieu, ou (b) pour se conformer à la législation en vigueur, à un processus judiciaire ou à une demande des autorités. Des sanctions peuvent être appliquées après l'enquête (voir la section 6 ci-après).
4.1 Notifications des suppressions ou des autres décisions d'exécution	Sans objet.
4.2 Mécanismes de recours en cas de suppression ou d'autres décisions d'exécution	Sans objet.
5. Moyens de détecter les contenus terroristes et extrémistes violents (par exemple, algorithmes de suivi, contenu généré par	Il appartient à la communauté Wikipedia de procéder au contrôle éditorial, et donc à la détection des contenus contraires aux règles de Wikipedia. Les lecteurs (utilisateurs de Wikipedia qui ne font pas de contribution) peuvent contacter l'Équipe bénévole de réponse pour signaler un problème sur un contenu en ligne. La Fondation Wikimedia déclare qu'elle n'assume aucune fonction éditoriale sur ses projets, y compris Wikipedia. Cela signifie qu'« en général », elle ne

les utilisateurs, évaluateurs humains (personnel), bases de données de partage d'empreintes numériques/d'adresses URL)	surveille pas et ne modifie pas le contenu des sites internet de ses projets (Wikimedia Foundation, 2019[178]). Les modérateurs de la communauté Wikipedia n'engendrent aucun coût pour la Fondation Wikimedia. Wikipedia n'est pas membre du GIFCT et ne participe pas au consortium de partage d'empreintes numériques.
6. Sanctions ou conséquences en cas de violation des Conditions générales d'utilisation ou des Règles de la communauté.	La communauté Wikipedia peut envoyer un avertissement, mener une enquête, supprimer des pages créées par des utilisateurs, bloquer ou exclure des utilisateurs qui enfreignent les règles de la communauté. La Fondation Wikimedia peut refuser, désactiver ou restreindre l'accès à la contribution d'un utilisateur qui enfreint ses Conditions générales d'utilisation, interdire à un utilisateur de contribuer à une page ou de modifier un contenu, ou bloque le compte ou l'accès d'un utilisateur qui a enfreint ses Conditions générales d'utilisation, et prendre des mesures judiciaires à son encontre (y compris envoyer un signalement aux autorités répressives).
7. Le service publie-t-il des rapports de transparence sur les contenus terroristes et extrémistes violents ?	Non. La Fondation Wikimedia publie des rapports de transparence (Wikimedia Foundation, s.d.[179]) qui comprennent les demandes de données des utilisateurs et les demandes de modification et de retrait de contenu, mais qui ne mentionnent pas précisément les contenus terroristes et extrémistes violents.
8. Quels informations ou types de données figurent-ils dans les rapports de transparence ?	Dans la partie Demandes des données des utilisateurs (Requests for user data), au paragraphe Divulgations d'urgence (Emergency disclosures), la Fondation Wikimedia publie le nombre de communications de données des utilisateurs liées à des menaces terroristes. Il ne correspond toutefois pas au nombre de contenus terroristes et extrémistes violents supprimés.
9. Méthodologies appliquées pour déterminer, calculer ou estimer les informations et données figurant dans les rapports de transparence.	Sans objet.
10. Fréquence de publication des rapports de transparence.	Sans objet.
11. Ce service a-t-il été utilisé pour mettre en ligne des contenus terroristes et extrémistes violents ?	Information inconnue.

Annexe C. **Définitions**

Les définitions suivantes sont utilisées aux fins du présent rapport :

Contenu : tout type d'information numérique servant à la diffusion de contenus terroristes et extrémistes violents, tels que des commentaires, images, vidéos, fichiers, posts, liens, dialogues en ligne, blogs ou messages.

Service de partage de contenus : tout service en ligne permettant le transfert, la transmission et la diffusion de contenus, sous quelque forme que ce soit, selon un rapport « un à un », un à une cible réduite, ou « un à *n* », indépendamment du caractère public, semi-privé ou privé du contenu. Tous les services pour lesquels un profil a été établi dans le présent rapport sont des services de partage de contenus en ligne.

Plateforme en ligne : service numérique qui facilite les interactions, par le biais de l'internet, entre au moins deux ensembles d'utilisateurs distincts mais interdépendants (personnes physiques ou morales).

Service de médias sociaux (ou de réseaux sociaux) : tout service en ligne permettant à des utilisateurs de créer un profil personnel public ou semi-public, de télécharger des contenus et d'accéder aux contenus partagés par d'autres utilisateurs, d'interagir et de nouer des relations avec d'autres utilisateurs, d'exprimer des avis et de partager des centres d'intérêt.

Utilisation de l'internet à des fins terroristes : utilisation de l'internet pour promouvoir des desseins terroristes (utilisation d'une application de messagerie pour coordonner un attentat terroriste, par exemple). La diffusion de contenus terroristes et extrémistes violents est un type d'utilisation de l'internet à des fins terroristes, qui peuvent être l'incitation à la violence, la radicalisation ou le recrutement.

Contenus terroristes et extrémistes violents : il n'existe pas de définition universellement admise du terrorisme ou de l'extrémisme violent, ni, par extension, des contenus terroristes et extrémistes violents. Le présent rapport s'appuie sur la terminologie utilisée dans l'Appel de Christchurch, et emploie cette expression pour désigner l'ensemble des contenus terroristes et extrémistes violents pour lesquels divers services de partage de contenus en ligne définissent des règles, mènent des activités de modération, prennent des décisions de retrait et, dans certains cas, établissent des rapports de transparence.

Références

4chan (s.d.), *'Advertise - 4chan'*, http://www.4chan.org/advertise (consulté le 31 août 2019). [63]

4chan (s.d.), *Frequently Asked Questions*, https://www.4channel.org/faq. [155]

Alexa (2019), *The top 500 sites on the web*, https://www.alexa.com/topsites/global;0. [68]

Alexander, J. (2019), *Verizon is selling Tumblr to WordPress' owner*, https://www.theverge.com/2019/8/12/20802639/tumblr-verizon-sold-wordpress-blogging-yahoo-adult-content. [135]

Amazon (s.d.), *Amazon.com Help: Law Enforcement Information Requests*, https://www.amazon.com/gp/help/customer/display.html?nodeId=GYSDRGWQ2C2CRYEF. [132]

Apple (s.d.), *Privacy - About Apple's Transparency Report*, https://www.apple.com/legal/transparency/about.html. [91]

Ardern, J. (2019), *Significant progress made on eliminating terrorist content online*, https://www.beehive.govt.nz/release/significant-progress-made-eliminating-terrorist-content-online. [197]

Arthur, R. (2019), *We Analyzed More Than 1 Million Comments on 4chan. Hate Speech There Has Spiked by 40% Since 2015*, https://www.vice.com/en_us/article/d3nbzy/we-analyzed-more-than-1-million-comments-on-4chan-hate-speech-there-has-spiked-by-40-since-2015. [156]

Artistic license (s.d.), *VK.com TakeDown Process*, https://www.artistic-license.org/takedowns/vk-com-takedown-process/. [199]

Automattic (s.d.), *Rapport de transparence*, https://transparency.automattic.com/. [171]

Baidu Inc. (2019), *Baidu, Inc. Files Its Annual Report on Form 20-F*, https://www.prnewswire.com/news-releases/baidu-inc-files-its-annual-report-on-form-20-f-300813428.html. [214]

Baidu Inc. (2019), *prnewswire.com*, https://www.prnewswire.com/news-releases/baidu-announces-third-quarter-2019-results-300953076.html. [203]

Baidu, Inc. (2018), *Form 20-F*, http://ir.iqiyi.com/static-files/83481f9b-238f-4841-9591-c0f9c817c7dc. [49]

Baidu, Inc. (2017), *Annual Report Pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934 for the fiscal year ended December 31, 2017*. [15]

Barnes, L. (2019), *One month after controversial adult-content purge, far-right pages are thriving on Tumblr*, https://thinkprogress.org/far-right-content-survived-tumblr-purge-36635e6aba4b/. [196]

Barr, J. (2016), *Does MySpace Have Any Distribution Juice Left for Publishers?*, https://adage.com/article/media/myspace-juice-left-publishers/303781. [64]

Bennett, C. (2019), *Extremism, George Washington University*, https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/EncryptedExtremism.pdf. [126]

Bicknell, Z. (2018), *What Video Platform Should I Use?*, https://www.theukdomain.uk/what-video-platform-should-i-use/. [52]

Birnbaum, E. (2019), *Social media giants restructure counterterrorism effort into independent group with staff*, https://thehill.com/policy/technology/462691-social-media-giants-restructure-counter-terrorism-effort-into-independent. [195]

Bradley MP, T. (2018), *Internet Safety Strategy green paper*, https://www.gov.uk/government/consultations/internet-safety-strategy-green-paper (consulté le 5 mai 2019). [43]

British Broadcasting Corporation (BBC) (2019), *Germany shooting: 2,200 people watched on Twitch*, https://www.bbc.com/news/technology-49998284. [134]

Carmen, A. (2015), *Filtered extremism: how ISIS supporters use Instagram*, https://www.theverge.com/2015/12/9/9879308/isis-instagram-islamic-state-social-media. [98]

Cheah, M. (2019), *Important updates to our content guidelines - Vimeo Blog*, https://vimeo.com/blog/post/important-updates-to-our-content-guidelines/. [122]

Cheng, J. (2014), *South Korea's KakaoTalk Adds 'Secret Mode'*, https://blogs.wsj.com/digits/2014/12/08/south-koreas-kakaotalk-adds-secret-mode/. [194]

Christchurch Call (2019), *Christchurch Call*, https://www.christchurchcall.com/call.html. [1]

Clicky, S. (2017), *Tackling Extremist Content on WordPress.com*, https://transparency.automattic.com/2017/12/06/tackling-extremist-content-on-wordpress-com/. [169]

Commission européenne (2019), *Countering illegal hate speech online #NoPlace4Hate*, https://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=54300%20 (consulté le 6 juin 2019). [33]

Commission européenne (2018), *Recommandation de la Commission sur les mesures destinées à lutter, de manière efficace, contre les contenus illicites en ligne*, https://ec.europa.eu/digital-single-market/en/news/commission-recommendation-measures-effectively-tackle-illegal-content-online. [35]

Commission européenne (2017), *Communication on Tackling Illegal Content Online - Towards an enhanced responsibility of online platforms*, https://ec.europa.eu/digital-single-market/en/news/communication-tackling-illegal-content-online-towards-enhanced-responsibility-online-platforms (consulté le 6 juin 2019). [34]

Conseil Constitutionnel (2020), *Décision n° 2020-801 DC du 18 juin 2020*, https://www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm. [39]

Conseil de Sécurité des Nations Unies (s.d.), *Liste récapitulative du Conseil de sécurité des Nations Unies*, https://www.un.org/securitycouncil/fr/content/un-sc-consolidated-list. [110]

Conway, M. (2019), « Disrupting Daesh: Measuring Takedown of Online Terrorist Material and Its Impacts, Studies in Conflict & Terrorism », vol. 42:/1-2, pp. 141-160, http://dx.doi.org/10.1080/1057610X.2018.1513984. [193]

Corfield, G. (2018), *The Register*, https://www.theregister.co.uk/2018/01/02/wechat_denial_user_surveillance/. [20]

Counter Extremism Project (2018), *Extremists & Online Propaganda*, https://www.counterextremism.com/sites/default/files/Extremists%20and%20Online%20Propaganda_040918.pdf. [209]

Counter Extremism Project (2018), *On Anniversary Of Barcelona Attacks, ISIS Continues Its Expansion*, https://www.counterextremism.com/press/anniversary-barcelona-attacks-isis-continues-its-expansion. [166]

Counter Terrorism Project (s.d.), *Extremists & Online Propaganda*, https://www.counterextremism.com/extremists-online-propaganda. [112]

Cox, J. (2019), *36 Days After Christchurch, Terrorist Attack Videos Are Still on Facebook*, https://www.vice.com/en_us/article/43jdbj/christchurch-attack-videos-still-on-facebook-instagram. [100]

Crawford, K. (2014), « What is a flag for? Social media reporting tools and the vocabulary of complaint », *New Media & Society*, vol. 18/3, pp. 410-428, https://journals.sagepub.com/doi/full/10.1177/1461444814543163. [13]

Creemers, R. (2018), *newamerica.org*, https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-cybersecurity-law-peoples-republic-china/. [17]

Datanyze (2019), *Market Share / Web Content Management Systems / October*, https://www.datanyze.com/market-share/wcms/october-market-share. [65]

Daum Kakao (s.d.), *Transparency Report, Kakao Privacy Policy*, http://privacy.daumkakao.com/en/transparence/report/request. [147]

Dearden, L. (2019), *Far-right extremists 'encouraged copycat terror attacks' after Christchurch mosque shootings*, https://www.independent.co.uk/news/uk/crime/far-right-terror-plots-uk-muslims-christchurch-attack-white-a9050511.html. [89]

Department of the Prime Minister and Cabinet (2019), *Australian Taskforce to Combat Terrorist and Extreme Violent Material Online*, https://www.pmc.gov.au/sites/default/files/publications/combat-terrorism-extreme-violent-material-online.pdf (consulté le 5 juin 2019). [32]

Deutscher Bundestag (2017), *Network Enforcement Act (Netzdurchsetzunggesetz, NetzDG)*, https://germanlawarchive.iuscomp.org/?p=1245 (consulté le 11 juin 2019). [41]

DeviantArt (s.d.), *What happens when my account is banned?*, https://www.deviantartsupport.com/en/article/what-happens-when-my-account-is-banned. [151]

DeviantArt (s.d.), *What is your policy around account suspensions?*, https://www.deviantartsupport.com/en/article/what-is-your-policy-around-account-suspensions. [150]

DeviantArt (s.d.), *What policy guidelines are there on comments, Journals, statuses, and general interactions?*, https://www.deviantartsupport.com/en/article/what-policy-guidelines-are-there-on-comments-journals-statuses-and-general-interactions. [149]

DeviantArt Media Kit (s.d.), *There's No Place Like DeviantArt*, https://deviantartads.com/. [62]

Dilger, D. (2015), *Another security manual recommends using Apple iMessage: this time, ISIS*, https://appleinsider.com/articles/15/11/21/another-security-manual-recommends-using-apple-imessage-this-time-isis-. [93]

Discord (2019), *Discord Transparency Report: Jan 1 — April 1*, https://blog.discordapp.com/discord-transparency-report-jan-1-april-1-4f288bf952c9?gi=e7efc9d05321. [146]

Dropbox (s.d.), *Security - Dropbox, Protecting your files*, https://www.dropbox.com/en_GB/security#files. [202]

Dropbox (s.d.), *Transparency Overview*, https://www.dropbox.com/en_GB/transparency. [165]

Dropbox (s.d.), *Who can see the stuff in my Dropbox account? Dropbox Help*, https://help.dropbox.com/accounts-billing/security/file-access. [164]

E&T editorial staff (2019), *E&T Engineering and Technology*, https://eandt.theiet.org/content/articles/2019/05/facebook-accused-of-exaggerating-success-in-tackling-extremism/. [210]

Elmer-Dewitt, P. (2019), *Information: Facebook's Messenger has overtaken Apple's iMessage*, https://247wallst.com/technology-3/2019/01/17/apple-facebook-messaging/. [46]

Engineering & Technology (2019), *Facebook accused of exaggerating success in tackling extremism*, https://eandt.theiet.org/content/articles/2019/05/facebook-accused-of-exaggerating-success-in-tackling-extremism/. [11]

Facebook (2020), *Community Standards Enforcement Report: Dangerous Organizations*, https://transparency.facebook.com/community-standards-enforcement#dangerous-organizations (consulté le 5 juin 2020). [7]

Facebook (2019), *Next Steps for the Global Internet Forum to Counter Terrorism - About Facebook*, https://about.fb.com/news/2019/09/next-steps-for-gifct/. [205]

Facebook (2018), *Hard Questions: What Are We Doing to Stay Ahead of Terrorists?*, https://about.fb.com/news/2018/11/staying-ahead-of-terrorists/. [72]

Facebook (2017), *Hard Questions: How We Counter Terrorism*, https://about.fb.com/news/2017/06/how-we-counter-terrorism/. [73]

Facebook (2018-2019), *Community Standards Enforcement, Terrorist Propaganda*, https://transparency.facebook.com/community-standards-enforcement#terrorist-propaganda. [74]

Facebook (s.d.), *Standards de la communauté 2. Individus et organismes dangereux*, https://www.facebook.com/communitystandards/dangerous_individuals_organizations. [69]

Facebook (s.d.), *Standards de la communauté 2. Individus et organismes dangereux*, https://www.facebook.com/communitystandards/dangerous_individuals_organizations. [70]

Facebook (s.d.), *Understanding the Community Standards Enforcement Report*, https://transparency.facebook.com/community-standards-enforcement/guide. [71]

Farrell, N. (2006), *Myspace is « terrorist recruiting ground »*, https://www.theinquirer.net/inquirer/news/1043174/myspace-is-terrorist-recruiting-ground. [157]

Financial Times (s.d.), *Businesses show no appetite for anti-terror AI tool*, https://www.ft.com/content/fda2d218-56fb-11e9-91f9-b6515a54c5b1. [213]

Financial Times (s.d.), *ISIS videos targeted by UK-funded artificial intelligence software*. [208]

Fisher-Birch, J. (2018), *Terror on Tumblr*, https://www.counterextremism.com/blog/terror-tumblr. [192]

Freedom House (2018), *France Country Report - Freedom on the Net 2018*, https://freedomhouse.org/country/france/freedom-net/2018 (consulté le 6 juin 2019). [37]

Frier, S. (2018), *Facebook Scans the Photos and Links You Send on Messenger*, https://www.bloomberg.com/news/articles/2018-04-04/facebook-scans-what-you-send-to-other-people-on-messenger-app. [90]

G20 (2019), *G20 Osaka Leaders' Statement on Preventing Exploitation of the Internet for Terrorism and Violent Extremism Conducive to Terrorism (VECT)*, https://dig.watch/instruments/g20-osaka-leaders-statement-preventing-exploitation-internet-terrorism-and-violent. [3]

G20 (2017), *The Hamburg G20 Leaders' Statement on Countering Terrorism*, https://www.g20germany.de/Content/DE/_Anlagen/G7_G20/2017-g20-statement-antiterror-en___blob=publicationFile&v=2.pdf (consulté le 15 janvier 2020). [2]

G7 (2019), *Résumé de la Présidence*, https://www.economie.gouv.fr/files/files/2019/G7/G7Num/Chairs_summary_version_finale_FR.pdf. [4]

Ghoshal, A. (2014), *GitHub, Vimeo and 30 more sites blocked in India over content from ISIS*. [191]

GIFCT (2019), *GIFCT Transparency Report*, https://gifct.org/transparency/. [27]

GIFCT (2019), *Membership Criteria*, https://www.gifct.org/members/. [14]

GIFCT (2019), *Next steps for the GIFCT*, https://gifct.org/press/next-steps-gifct/. [30]

GIFCT (2019), *Progress for the Independent GIFCT*, https://gifct.org/press/progress-independent-gifct/. [29]

GIFCT (2017), *Global Internet Forum to Counter Terrorism: Evolving an Institution*, https://gifct.org/about/. [26]

GIFCT (2017), *Joint Tech Innovation*, https://gifct.org/joint-tech-innovation/. [28]

Google (2019), *Conditions d'utilisation supplémentaires de Google Drive*, https://www.google.com/intl/fr/drive/terms-of-service/. [160]

Google (2019), *Rapport de transparence*, https://transparencyreport.google.com/youtube-policy/featured-policies/violent-extremism?hl=fr_FR. [190]

Google (2010), *Tools to visualize access to information*, https://publicpolicy.googleblog.com/2010/09/tools-to-visualize-access-to.html. [9]

Google (s.d.), *Aide Éditeurs Docs - Demander l'examen d'un rapport pour infraction au règlement*, https://support.google.com/docs/answer/2463328?hl=fr&ref_topic=1360897. [159]

Google (s.d.), *Aide Éditeurs Docs - Règlement du programme concernant l'utilisation abusive et mise en application*, https://support.google.com/docs/answer/148505?visit_id=637064013896463652-1393240150&rd=1. [158]

Google (s.d.), *Aide Éditeurs Docs - Signaler une infraction*, https://support.google.com/docs/answer/2463296?hl=fr&ref_topic=1360897. [162]

Google (2010-2019), *Demandes gouvernementales de suppression de contenu*, https://transparencyreport.google.com/government-removals/overview?hl=fr_FR. [87]

Google (s.d.), *Transparence des informations*, https://transparencyreport.google.com/about?hl=fr_FR. [88]

Google (s.d.), *Transparence des informations*, https://transparencyreport.google.com/?hl=fr_FR. [85]

Google, YouTube (s.d.), *Aide YouTube - Fonctionnalités limitées pour certaines vidéos*, https://support.google.com/youtube/answer/7458465. [84]

Google, YouTube (s.d.), *Community Guidelines strike basics - YouTube Help*, https://support.google.com/youtube/answer/2802032. [83]

Google, YouTube (s.d.), *Sélection de règles : Contenu extrémiste violent*, https://transparencyreport.google.com/youtube-policy/featured-policies/violent-extremism?hl=fr_FR&policy_removals=period:Y2019Q2&lu=policy_removals. [86]

Google, Youtube (2020), *Activer ou désactiver le mode restreint*, https://support.google.com/youtube/answer/174084?hl=fr. [77]

Google, Youtube (2020), *Faire appel des actions pour non-respect du règlement de la communauté*, https://support.google.com/youtube/answer/185111?hl=fr. [76]

Google, Youtube (2020), *Signaler un contenu inapproprié*, https://support.google.com/youtube/answer/2802027?hl=fr. [78]

Google, Youtube (2020), *YouTube Trusted Flagger program*, https://support.google.com/youtube/answer/7554338?&ref_topic=2803138. [79]

Google, Youtube (s.d.), *Sélection de règles : Incitation à la haine*, https://transparencyreport.google.com/youtube-policy/featured-policies/hate-speech?hl=fr_FR. [82]

Google/ YouTube (2020), *Règles concernant les organisations criminelles violentes*, https://support.google.com/youtube/answer/9229472?hl=fr&ref_topic=9282436. [75]

Gouvernement de l'Australie, Federal Register of Legislation (2019), *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*, https://www.legislation.gov.au/Details/C2019A00038. [31]

Gouvernement du Royaume-Uni (2019), *Online Harms White Paper*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf (consulté le 4 juin 2019). [44]

Gouvernement français (2018), *Rapport visant à renforcer la lutte contre le racisme et l'antisémitisme sur internet*, https://www.gouvernement.fr/partage/10528-rapport-visant-a-renforcer-la-lutte-contre-le-racisme-et-l-antisemitisme-sur-internet (consulté le 6 juin 2019). [38]

Hatmaker, T. (2019), *This led to Reddit administrators banning the entire community in question from the site*, https://techcrunch.com/2019/03/15/reddit-watchpeopledie-subreddit-gore/. [105]

Hayden, M. (2019), *Far-Right Extremists Are Calling for Terrorism on the Messaging App Telegram*, https://www.splcenter.org/hatewatch/2019/06/27/far-right-extremists-are-calling-terrorism-messaging-app-telegram. [127]

Hayden, M. (2019), *Mysterious Neo-Nazi Advocated Terrorism for Six Years Before Disappearance*, https://www.splcenter.org/hatewatch/2019/05/21/mysterious-neo-nazi-advocated-terrorism-six-years-disappearance. [152]

Home Office, A. (2018), *New technology revealed to help fight terrorist content online*, https://www.gov.uk/government/news/new-technology-revealed-to-help-fight-terrorist-content-online. [189]

Huang, F. (2018), *China's Most Popular App Is Full of Hate*, https://foreignpolicy.com/2018/11/27/chinas-most-popular-app-is-full-of-hate/. [96]

Hymas, C. (2019), *Isil extremists using Instagram to promote jihad and incite support for terror attacks on the West*, https://www.telegraph.co.uk/news/2019/05/11/isil-extremists-using-instagram-promote-jihad-incite-support/. [99]

Instagram (2019), *Changes to Our Account Disable Policy*, https://instagram-press.com/blog/2019/07/18/changes-to-our-account-disable-policy/. [97]

Investopedia (2019), *Monthly Active User (MAU) Definition*, https://www.investopedia.com/terms/m/monthly-active-user-mau.asp. [206]

Iqbal, M. (2019), *Twitch Revenue and Usage Statistics (2019)*, https://www.businessofapps.com/data/twitch-statistics/. [56]

ISDGlobal (s.d.), *Powering solutions to extremism and polarisation*, https://www.isdglobal.org/. [81]

Kallas, P. (2019), *Top 15 Most Popular Social Networking Sites and Apps [2020] @ Dreamgrow*, https://www.dreamgrow.com/top-15-most-popular-social-networking-sites/. [55]

Katz, R. (2019), *A Growing Frontier for Terrorist Groups: Unsuspecting Chat Apps*, https://www.wired.com/story/terrorist-groups-prey-on-unsuspecting-chat-apps/. [119]

Katz, R. (2018), *To Curb Terrorist Propaganda Online, Look to YouTube. No, Really*, https://www.wired.com/story/to-curb-terrorist-propaganda-online-look-to-youtube-no-really/. [163]

Kemp, S. (2019), *Data 2019: Global Digital Overview*, https://datareportal.com/reports/digital-2019-global-digital-overview. [47]

Kemp, S. (2019), *The Global State of Digital in October 2019*, https://wearesocial.com/blog/2019/10/the-global-state-of-digital-in-october-2019. [45]

Kenny, K. (2019), *How can upcoming social media efforts be 'global' if they ignore Asia?*, https://www.stuff.co.nz/national/christchurch-shooting/112284082/how-can-upcoming-social-media-efforts-be-global-if-they-ignore-asia. [95]

Kinsta (2011-2019), *Wordpress Market Share Statistics (2011-2019)*, https://kinsta.com/wordpress-market-share/. [66]

Kitsune, L. (2017), *New Notifications and Reporting Updates by Lauren Kitsune on DeviantArt*, https://www.deviantart.com/laurenkitsune/journal/New-Notifications-and-Reporting-Updates-706864447. [148]

Knockel, J. (2018), *(Can't) Picture This, An Analysis of Image Filtering on WeChat Moments*, https://citizenlab.ca/2018/08/cant-picture-this-an-analysis-of-image-filtering-on-wechat-moments/. [16]

Knockell, J. (2015), *Every Rose Has Its Thorn: Censorship and Surveillance on Social VideoPlatforms in China*, https://www.usenix.org/system/files/conference/foci15/foci15-paper-knockel.pdf. [140]

Korea Legislation Research Institute, K. (2016), *ACT ON COUNTER-TERRORISM FOR THE PROTECTION OF CITIZENS AND PUBLIC SECURITY*, http://elaw.klri.re.kr/eng_service/lawView.do?hseq=38450&lang=ENG%20 (consulté le 11 June 2019). [188]

Lakomy, M. (2017), « Cracks in the Online "Caliphate": How the Islamic State is Losing Ground in the Battle for Cyberspace. », *Perspectives On Terrorism*, vol. 11/3, http://www.terrorismanalysts.com/pt/index.php/pot/article/view/607. [187]

Lange, D. (2017), *Quora's Tolerance Of Terror Support*, https://www.israellycool.com/2017/05/22/quoras-tolerance-of-terror-support/. [114]

Leisegang, D. (2017), *No freedom to hate: Germany's new law against online incitement*, https://www.eurozine.com/no-freedom-to-hate-germanys-new-law-on-online-incitement/ (consulté le 6 juin 2019). [40]

Liao, S. (2018), *Discord shuts down more neo-Nazi, alt-right servers*, https://www.theverge.com/2018/2/28/17062554/discord-alt-right-neo-nazi-white-supremacy-atomwaffen. [145]

LINE (2019), *LINE Content Moderation Report*. [128]

LINE (s.d.), *Help Center*, https://help.line.me/line/android/categoryId/20000132/3/pc?lang=en. [129]

LinkedIn (s.d.), *Our Transparency Report*, https://about.linkedin.com/transparency. [204]

Manileve, V. (2016), *The Problem With Snapchat's Coverage of the Terror in Nice*, https://slate.com/technology/2016/07/did-snapchat-show-its-users-too-much-from-the-tragedy-in-nice.html. [186]

Marketing Land (2 018), *Quora Introduces Broad Targeting, Says Audience Hits 300 Million Monthly Users*, https://marketingland.com/quora-introduces-broad-targeting-says-audience-hits-300-million-monthly-users-248261. [51]

Marshall, C. (2019), *Twitch suspends streaming for new users as it fights off Artifact trolls*, https://www.polygon.com/2019/5/28/18643198/twitch-artifact-section-stream-suspended. [133]

Medium (2015), *Medium's Transparency Report (2014)*, https://medium.com/transparency-report/mediums-transparency-report-438fe06936ff. [142]

Medium (s.d.), *Controversial, Suspect and Extreme Content, Medium Help Center*, https://help.medium.com/hc/en-us/articles/360018182453. [141]

Meetup (2019), *Conditions générales de service*, https://help.meetup.com/hc/fr-fr/articles/360027447252-Conditions-g%C3%A9n%C3%A9rales-de-service. [153]

Meetup (2017), *Introducing Meetup's Inaugural Transparency Report*, http://blog.meetup.com/inaugural-transparency-report/. [154]

Microsoft (2019), *Contents Removals Request Report, Microsoft CSR*, https://www.microsoft.com/en-us/corporate-responsibility/content-removal-requests-report. [111]

Microsoft (2016), *Microsoft's approach to terrorist content online, Microsoft on the Issues*, https://blogs.microsoft.com/on-the-issues/2016/05/20/microsofts-approach-terrorist-content-online/#sm.000del1ea19zbe4duja1ve96fcc1l. [108]

Microsoft (s.d.), *Comment OneDrive protège vos données dans le cloud*, https://support.microsoft.com/fr-fr/office/comment-onedrive-prot%c3%a8ge-vos-donn%c3%a9es-dans-le-cloud-23c6ea94-3608-48d7-8bf0-80e142edd1e1?ui=fr-FR&rs=fr-FR&ad=FR. [211]

Miller, J. (2014), *Can Iraqi militants be kept off social media sites?*. [130]

Moign Khawaja, S. (2019), « Disrupting Daesh: Measuring Takedown of Online Terrorist Material and Its Impacts », *Studies in Conflict & Terrorism*, vol. 42/1-2. [185]

Murphy, N. (2019), *Reddit's 2019 Year in Review*, https://redditblog.com/2019/12/04/reddits-2019-year-in-review/. [50]

New America (s.d.), *Case Study 3 - Transparency Reporting*, https://www.newamerica.org/in-depth/getting-internet-companies-do-right-thing/case-study-3-transparency-reporting/. [10]

New America (s.d.), *The Transparency Reporting Toolkit: Content Takedown Reporting*, https://www.newamerica.org/oti/reports/transparency-reporting-toolkit-content-takedown-reporting/introduction-and-executive-summary/. [8]

Odnoklassniki (s.d.), *Help Centre*, https://ok.ru/help/54/367. [143]

OFCOM (2019), *Use of AI in Online Content Moderation – Report Produced on Behalf of Ofcom*, https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf. [12]

Office fédéral allemand de la justice (2019), *Federal Office of Justice Issues Fine against Facebook*, https://www.bundesjustizamt.de/DE/Presse/Archiv/2019/20190702_EN.html. [42]

Parlement européen (2019), *Résolution législative du Parlement européen du 17 avril 2019 sur la proposition de règlement du Parlement européen et du Conseil relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne (COM(2018)0640 – C8-0405/2018 – 2018/0331(COD))*, https://www.europarl.europa.eu/doceo/document/TA-8-2019-0421_FR.html. [36]

Pavel, D. (2018), *200,000,000 Monthly Active Users*, https://telegram.org/blog/200-million. [54]

Pew Research Center (2016), *Wikipedia at 15: Millions of readers in scores of languages*, https://www.pewresearch.org/fact-tank/2016/01/14/wikipedia-at-15/. [67]

Pinterest (2019), *Transparency Report - Pinterest help*, https://help.pinterest.com/en-gb/article/transparency-report. [121]

Pinterest (s.d.), *Suspension de compte*, https://help.pinterest.com/fr/article/account-suspension. [120]

Powell, B. (2019), *Encrypted Extremism - Inside the English-Speaking Islamic State Ecosystem on Telegram*. [144]

Quora (s.d.), *How does Quora Moderation make decisions about edit-blocks and bans? How does someone appeal this decision?*, https://www.quora.com/How-does-Quora-Moderation-make-decisions-about-edit-blocks-and-bans-How-does-someone-appeal-this-decision. [113]

Rakuten Viber (s.d.), *Viber Encryption Overview*, https://www.viber.com/app/uploads/viber-encryption-overview.pdf. [200]

Reddit Inc. (2018), *Transparency Report 2018*, https://www.redditinc.com/policies/transparency-report-2018. [104]

Reddit Inc. (2017), *Moderator Guidelines for Healthy Communities*, https://www.redditinc.com/policies/moderator-guidelines. [102]

Reddit Inc. (s.d.), *AutoModerator*, https://mods.reddithelp.com/hc/en-us/articles/360002561632-AutoModerator. [103]

Reddit Inc. (s.d.), *Quarantined Subreddits*, https://www.reddithelp.com/en/categories/rules-reporting/account-and-community-restrictions/quarantined-subreddits. [101]

Roberts, S. (2019), *Behind the Screen: Content Moderation in the Shadows of Social Media*, Yale University Press, http://dx.doi.org/9780300235883. [184]

Roberts, S. (2017), *Content Moderation*, https://escholarship.org/uc/item/7371c1hf. [5]

Roettgers, J. (2019), *Facebook Gets One-Strike Policy for Live Streaming, Twitter and Twitch May Be Next*, https://variety.com/2019/digital/news/facebook-live-streaming-rules-1203215794/. [183]

Rosenthal, M. (dir. pub.) (2014), *Vkontakte, a Russian social network, is hosting ISIS accounts that were kicked off of Facebook and Twitter*, https://www.pri.org/stories/2014-09-12/isis-internet-army-has-found-safe-haven-russian-social-networks-now. [139]

Ruan, L. (2016), *One App, Two Systems, How WeChat uses one censorship policy in China and another internationally*, https://citizenlab.ca/2016/11/wechat-china-censorship-one-app-two-systems/. [21]

Saima, S. (2019), *Finally: Snapchat comes up with end-to-end encryption to secures users conversations and data*, https://www.digitalinformationworld.com/2019/01/snapchat-end-to-end-encryption-users-media-messages.html#. [182]

Santa Clara University's High Tech Law Institute (s.d.), *The Santa Clara Principles On Transparency and Accountability in Content Moderation*, https://santaclaraprinciples.org/. [115]

Site Intelligence Group Enterprise (2018), *IS-linked Media Group Makes Foray onto Viber Messenger - Dark Web and Cyber Security*, https://ent.siteintelgroup.com/Dark-Web-and-Cyber-Security/is-linked-media-group-makes-foray-onto-viber-messenger.html. [118]

Snap Inc (s.d.), *Snaps et Chats*, https://www.snap.com/fr-FR/privacy/privacy-by-product. [201]

Snap Inc. (2019), *Rapport sur la transparence (1 janvier 2020 – 30 juin 2020)*, https://www.snap.com/fr-FR/privacy/transparency. [117]

Snap Inc. (s.d.), *Signaler un problème de sécurité*, https://www.snap.com/fr-FR/safety/safety-reporting. [116]

Solsman, J. (2018), *« Smule May Be the Biggest Music App You Haven't Heard Of »*, https://www.cnet.com/news/smule-is-the-biggest-music-app-you-never-heard-of/. [60]

START (National Consortium for the Study of Terrorism and Responses to Terrorism) (2018), *The Use of Social Media by United States Extremists*, https://www.start.umd.edu/pubs/START_PIRUS_UseOfSocialMediaByUSExtremists_ResearchBrief_July2018.pdf. [109]

Statista (2019), *Number of global monthly active Kakaotalk users from 1st quarter 2013 to 1st quarter 2019*, https://www.statista.com/statistics/278846/kakaotalk-monthly-active-users-mau/. [61]

Tech Against Terrorism (2019), *Analysis: ISIS use of smaller platforms and the DWeb to share terrorist content*, https://www.techagainstterrorism.org/2019/04/29/analysis-isis-use-of-smaller-platforms-and-the-dweb-to-share-terrorist-content-april-2019/. [6]

Telegram (s.d.), *ISIS Watch*, https://telegram.me/ISISwatch. [125]

Telegram (s.d.), *Telegram Privacy Policy*, https://telegram.org/privacy?setln=fr. [124]

Tencent (s.d.), *Agreement on Software License and Service of Tencent Weixin*, https://weixin.qq.com/cgi-bin/readtemplate?lang=en&t=weixin_agreement&s=default&cc=CN. [94]

The Associated Press (2014), *China tightens social media control, tells South Korea some used for terror information*, http://www.vancouversun.com/China+tightens+social+media+control+tells+South+Korea+some+used+terror+information/10097184/story.html. [212]

The International Centre for the Study of Radicalisation (ICSR) (2020), *ICSR info*, https://icsr.info/. [80]

The Tech Against Terrorism team, L. (2019), *techagainstterrorism.org*, https://www.techagainstterrorism.org/2019/04/29/analysis-isis-use-of-smaller-platforms-and-the-dweb-to-share-terrorist-content-april-2019/. [181]

The Washington Post (2019), *Tiktok's Beijing roots fuel censorship suspicion as it builds a huge U.S. audience*, https://www.washingtonpost.com/technology/2019/09/15/tiktoks-beijing-roots-fuel-censorship-suspicion-it-builds-huge-us-audience/. [22]

TikTok (2019), *Our committment to our users and the TikTok experience*, https://newsroom.tiktok.com/en-us/our-commitment-to-our-users-and-the-tik-tok-experience. [24]

TikTok (2019), *TikTok Transparency Report*, https://www.tiktok.com/safety/resources/transparency-report. [25]

Titcomb, J. (2017), *Why Google is reading your Docs*, https://www.telegraph.co.uk/technology/2017/11/01/google-reading-docs/. [161]

Twitch (s.d.), *Utilisation d'AutoMod*, https://help.twitch.tv/s/article/how-to-use-automod?language=fr. [131]

Twitter (2019), *Twittter Rules enforcement*, https://transparency.twitter.com/en/twitter-rules-enforcement.html. [107]

Twitter (s.d.), *Notre gamme d'options pour l'application de nos politiques*, https://help.twitter.com/fr/rules-and-policies/enforcement-options. [106]

US Treasury (2020), *OFFICE OF FOREIGN ASSETS CONTROL - Specially Designated Nationals and Blocked Persons List*, https://www.treasury.gov/ofac/downloads/sdnlist.pdf. [167]

Verizon Media (2019), *Rapport de transparence*, https://www.verizonmedia.com/transparency/index.html?guce_referrer=aHR0cHM6Ly90cmFuc3BhcmVuY3kub2F0aC5jb20vaW5kZXguaHRtbD9ndWNlX3JlZmVycmVyPWFIUjBjSE02THk5M2QzY3VkSFZ0WW14eUxtTnZiUzgmZ3VjZV9yZWZlcnJlcl9zaWc9QVFBQUFKazduZ3VNWS04dHhtNG9hWFM3TUlkNkxlUWxkMEZ5. [136]

Vimeo (s.d.), *Comment Vimeo gère-t-il les contenus à caractère violent ? - Centre d'aide*, https://vimeo.zendesk.com/hc/fr/articles/224822427-Comment-Vimeo-g%C3%A8re-t-il-les-contenus-%C3%A0-caract%C3%A8re-violent-. [123]

Vincent, B. (2019), *Discord Celebrates Four Years With 250 Million Users*, https://variety.com/2019/gaming/news/discord-anniversary-250-million-users-1203213244/. [59]

Wang, M. (2019), *Wechatscope*, https://advox.globalvoices.org/2019/02/11/censored-on-wechat-a-year-of-content-removals-on-chinas-most-powerful-social-media-platform/. [19]

Weimann, G. (2014), *New Terrorism and New Media*, Commons Lab of the Woodrow Wilson International Center for Scholars, https://www.wilsoncenter.org/sites/default/files/new_terrorism_v3_1.pdf. [137]

Wickey, W. (2018), *Should You Use Medium As Your Business Blog Platform? [2019 Update]*, https://medium.com/crowdbotics/medium-business-blog-platform-b8b8faa2d430. [58]

Wikimedia Foundation (2019), *Conditions d'utilisation - Wikimedia Foundation - Governance Wiki*, https://foundation.wikimedia.org/wiki/Terms_of_Use/fr. [178]

Wikimedia Foundation (s.d.), *Transparency report*, https://transparency.wikimedia.org/. [179]

Wikipedia (2020), *Administration - Wikipedia*, https://en.wikipedia.org/wiki/Wikipedia:Administration#Human_and_legal_administration. [173]

Wikipedia (2020), *Deletion process - Wikipedia*, https://en.wikipedia.org/wiki/Wikipedia:Deletion_process. [176]

Wikipedia (2020), *What Wikipedia is not - Wikipedia*. [175]

Wikipedia (2020), *Wikipédia :Critères de suppression immédiate*, https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Crit%C3%A8res_de_suppression_imm%C3%A9diate. [177]

Wikipedia (2019), *Core Content Policies - Wikipedia*, https://en.wikipedia.org/wiki/Wikipedia:Core_content_policies. [174]

Wikipedia (2019), *Wikipédia:Vérificateur d'adresses IP*, https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:V%C3%A9rificateur_d%27adresses_IP. [172]

Wikipedia (2 020), *Wikipédia:Masqueur de modifications*, https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Masqueur_de_modifications. [198]

Wired, L. (2019), *TikTok, under scrutiny, distances itself from China*, https://www.wired.com/story/tiktok-under-scrutiny-china/. [23]

Wohn, D. (2019), *Volunteer Moderators in Twitch Micro Communities: How They Get Involved, the Roles They Play, and the Emotional Labor They Experience*, http://dx.doi.org/10.1145/3 290 605,3300390. [180]

WordPress (s.d.), *Legal Guidelines - Support*, https://en.support.wordpress.com/report-blogs/legal-guidelines/. [207]

WordPress (s.d.), *Terrorist Activity - Support - Word Press.com*, https://en.support.wordpress.com/terrorist-activity/. [168]

WordPress.com (s.d.), *Contenu et sites suspendus*, https://wordpress.com/fr/support/blogs-suspendus/. [170]

Yang, S. (2019), *A War Between Two Chinese Internet Giants: Baidu and ByteDance*, https://equalocean.com/analysis/2019050511004. [57]

Yoo, E. (2018), *Huoshan latest video platform to clean up vulgar content*, https://technode.com/2018/04/13/huoshan-clean-up/. [138]

Youku Tudou Inc. (NYSE: YOKU) (s.d.), *Youku Tudou Inc. (NYSE: YOKU), About us - 优酷视频*, https://c.youku.com/abouteg/youtu. [48]

YY Inc. - IR Site (2019), *YY Reports First Quarter 2019 Unaudited Financial Results*, http://ir.yy.com/news-releases/news-release-details/yy-reports-first-quarter-2019-unaudited-financial-results. [53]

Zetter, K. (2015), *Security Manual Reveals the OPSEC Advice ISIS Gives Recruits*, https://www.wired.com/2015/11/isis-opsec-encryption-manuals-reveal-terrorist-group-security-protocols/. [92]

Zhang, Y. (2018), *Global Times*, http://www.globaltimes.cn/content/1098173.shtml. [18]

Notes

[1] Il n'existe pas de définition universellement admise du terrorisme ou de l'extrémisme violent, ni, par extension, des contenus terroristes et extrémistes violents. Le présent rapport ne prétend pas en fournir une, et utilise ces termes selon l'usage qui en est fait dans l'Appel de Christchurch.

[2] L'utilisation de l'internet à des fins terroristes peut être appréhendée comme le problème général de l'usage de l'internet pour promouvoir des desseins terroristes et extrémistes violents, sous quelque forme que ce soit ; en revanche, la diffusion de contenus terroristes et extrémistes violents est un type d'utilisation de l'internet à des fins terroristes, généralement dans le but d'inciter à la violence, à la radicalisation et au recrutement d'individus.

[3] Voir à l'Annexe C les définitions d'autres termes utilisés dans ce rapport.

[4] Voir la colonne « Type de service » des tableaux figurant à l'Annexe A.

[5] « Le nombre d'utilisateurs actifs mensuels renseigne sur la santé globale d'une entreprise en ligne et est utilisé pour le calcul d'autres indicateurs relatifs aux sites web. Il est également utile pour mesurer l'efficacité des campagnes de marketing d'une entreprise, ainsi que l'expérience des clients et prospects. Les investisseurs spécialisés dans le secteur des médias sociaux surveillent de près ce chiffre lorsque les entreprises le publient, car cet [indicateur de performance clé] peut influer sur le cours de bourse d'une entreprise de médias sociaux. » (Investopedia, 2019[206])

[6] Les informations issues des médias et autres sources librement accessibles ont en revanche été utilisées pour la section 10 de chaque profil (voir Annexe B), principalement parce que les documents constitutifs des services mentionnent rarement des incidents concrets liés à l'exploitation de leurs technologies à des fins terroristes et extrémistes violentes. En tout état de cause, ces sources, lorsqu'elles sont utilisées, sont dûment référencées dans les notes de bas de page des profils concernés.

[7] Voir la section 1 des profils suivants : Facebook, YouTube, WhatsApp, iMessage/FaceTime, WeChat, Instagram, QQ, Youku Tudou, QZone, TikTok, Weibo, iQIYI, Reddit, Twitter, Douban, LinkedIn, Baidu Tieba, Skype, Quora, Snapchat, Viber, Pinterest, Vimeo, IMO, LINE, Ask.fm, Twitch, YY Live, Xigua, Tumblr, Flickr, Huoshan, VK, Medium, Haokan, Odnoklassniki, Discord, Smule, KaKaoTalk, DeviantArt, Meetup, MySpace, Google Drive, Dropbox, OneDrive et WordPress.com.

[8] Voir la section 1 des profils suivants : Facebook, YouTube, TikTok, Twitter et Google Drive. On peut considérer que Microsoft (Skype et OneDrive) en fait partie, bien que l'entreprise ne donne pas de définition de l'extrémisme violent et ne fournisse pas d'exemple.

[9] Voir la section 1 des profils suivants : Instagram, Youku Tudou, iQIYI, LinkedIn, Baidu Tieba, Skype, Quora, Snapchat, Pinterest, Ask.fm, Xigua, Tumblr, Flickr, Huoshan, Haokan, Meetup, Dropbox, Microsoft OneDrive et Wordpress.com.

[10] Voir la section 1 des profils suivants : WeChat, Instagram, QQ, Youku Tudou, iQIYI, Douban, LinkedIn, Baidu Tieba, Vimeo, Twitch, Medium, Odnoklassniki, KaKaoTalk, Meetup et MySpace.

[11] Voir la section 1 des profils suivants : WhatsApp, iMessage/FaceTime, QZone, Weibo, Reddit, Viber, IMO, Telegram, LINE, VK, YY Live, Discord, Smule, DeviantArt, 4chan et Wikipedia.

[12] Voir la section 7 du profil de Facebook.

[13] Voir la section 7 du profil de YouTube.

[14] Des critiques pourraient être formulées à l'encontre de cette approche, certains gouvernements qualifiant les groupes politiques rivaux de terroristes. Quora, à l'inverse, se réfère à la liste des organisations terroristes étrangères, établie par le Département d'État des États-Unis. Voir la section 1 du profil de Quora. WordPress.com, qui se fonde sur la liste des *Special Designated Nationals* du Bureau du Département du Trésor chargé du contrôle des avoirs étrangers, suit une approche similaire. Voir la section 1 du profil de WordPress.com.

[15] Voir la section 1 des profils de Skype et OneDrive.

[16] Par exemple, une étude réalisée à partir d'un ensemble de données comprenant 722 comptes pro-EI et 451 autres comptes de djihadistes sur Twitter a révélé que les autres djihadistes sont parvenus à publier sur leurs comptes six fois plus de contenus et comptaient 13 fois plus d'abonnés sur Twitter que les pro-EI. De même, alors que 25 % des comptes pro-EI ont été suspendus dans un délai de cinq jours après leur création, le taux de retrait était inférieur à 1 % pour les comptes des autres djihadistes. Les auteurs en ont conclu que sur Twitter, tous les djihadistes ne sont pas soumis aux mêmes mesures que l'EI (Moign Khawaja, 2019[185]).

[17] Pour en savoir plus sur la genèse de l'établissement des rapports de transparence par les entreprises de l'internet, voir : https://www.newamerica.org/in-depth/getting-internet-companies-do-right-thing/case-study-3-transparency-reporting/.

[18] Voir la section 7 des profils suivants : Facebook, YouTube, Apple, Instagram, TikTok, Reddit, Twitter, LinkedIn, Skype, Snapchat, Pinterest, LINE, Twitch, Tumblr, Medium, Discord, KaKaoTalk, Meetup, Google Drive, Dropbox, OneDrive, WordPress.com et Wikipedia. Il convient de noter que Facebook Messenger n'est pas comptabilisé séparément ici car Facebook ne publie pas de rapport de transparence distinct pour Messenger.

[19] Voir les sections 7 à 9 des profils suivants : Facebook, YouTube, Instagram, Twitter et WordPress.com. Il convient de souligner que Discord a indiqué avoir reçu des signalements liés à la diffusion en direct de l'attentat de Christchurch, mais qu'il s'agit là de la seule information communiquée sur des contenus terroristes et extrémistes violents. Voir la section 8 du profil de Discord. Wikimedia communique le nombre de « divulgations d'urgence » de données d'utilisateurs en cas de menace terroriste, mais cela ne s'apparente pas à des retraits de contenus terroristes et extrémistes violents. Voir la section 8 du profil de Wikipedia. Par ailleurs, LinkedIn fait rapport sur les infractions à ses règles relatives aux contenus « violents ou choquants », mais outre les contenus qui profèrent des menaces ou promeuvent le terrorisme ou la violence, et ceux qui sont d'une violence extrême, la plateforme inclut également une catégorie « autre activité criminelle » et les contenus destinés à choquer ou humilier des individus. Le champ est donc plus large que les seuls contenus terroristes et extrémistes violents. Il convient de noter que Facebook Messenger n'est pas comptabilisé séparément ici car Facebook ne publie pas de rapport de transparence distinct pour Messenger.

[20] Des sujets comme le changement climatique, les minéraux des conflits, ou le harcèlement, l'exploitation et les atteintes sexuels en sont des exemples.

[21] Voir la section 8 du profil de WordPress.com.

[22] Voir la section 8 du profil de Twitter.

[23] Voir la section 8 du profil de YouTube.

[24] Voir la section 8 du profil de Facebook.

[25] Voir la section 8 du profil d'Instagram.

[26] Voir la section 8 du profil de YouTube.

[27] Voir la section 8 du profil de YouTube.

[28] Voir la section 8 du profil de Facebook.

[29] Voir la section 8 du profil de Twitter.

[30] Voir les sections 8 et 9 du profil de Facebook.

[31] YouTube a indiqué que près de 90 % des vidéos mises en ligne en septembre 2018 et supprimées pour non-respect du règlement sur l'extrémisme violent n'ont pas atteint 10 vues (voir https://transparencyreport.google.com/youtube-policy/featured-policies/violent-extremism?hl=fr). Il n'indique pas en revanche la méthodologie utilisée pour arriver à ce chiffre.

[32] Voir la section 8 des profils de Facebook, YouTube et Twitter.

[33] Cela va dans le sens des propositions du groupe de réflexion australien. Voir la partie consacrée à l'Australie, à la section 4 du présent rapport.

[34] Par exemple, le mécanisme de signalement de Twitter a été détourné par des suprémacistes blancs pour obtenir la clôture de comptes de féministes qui utilisaient le mot-dièse #solidarityisforwhitewomen (Crawford, 2014[13]).

[35] Voir la section 5 des profils suivants : Facebook, YouTube, WhatsApp, Facebook Messener, iMessage/FaceTime, Instagram, TikTok, Weibo, Reddit, Twitter, LinkedIn, Baidu Tieba, Skype, Quora, Snapchat, Viber, Pinterest, Vimeo, Telegram, LINE, Ask.fm, Xigua, Tumblr, Flickr, Houshan, VK, Medium, Odnoklassniki, Discord, Smule, KaKaoTalk, DeviantArt, Meetup, 4chan, MySpace, Google Drive, Dropbox, OneDrive, WordPress.com et Wikipedia.

[36] Voir les sections 4 et 5 des profils suivants : Reddit, Viber, Twitch, Flickr, VK, Odnoklassniki, KaKaoTalk, DeviantArt, 4chan et Wikipedia.

[37] L'expression « au moins » est indiquée ici car il n'est pas possible de déterminer, à partir des informations librement accessibles, le type d'activité et de processus mis en place par les services pour veiller au respect de leurs conditions générales d'utilisation et autres documents constitutifs. Voir par exemple la section 5 des profils suivants : QQ, Youku Tudou, QZone TikTok, Weibo, iQIYI, Douban, Baidu Tieba, YY Live, Xigua, Huoshan et Haokan.

[38] Voir la section 5 des profils suivants : Facebook, YouTube, WhatsApp, Facebook Messenger, WeChat, Instagram (membre du Hash Sharing Consortium), TikTok, Reddit (membre du Hash Sharing Consortium), Twitter, LinkedIn (membre du Hash Sharing Consortium), Skype (membre indirect du GIFCT via Microsoft), Snapchat (membre du Hash Sharing Consortium), Pinterest (membre du GIFCT), LINE, Ask.fm (membre du Hash Sharing Consortium), Twitch (membre indirect du GIFCT via Amazon), VK, YY Live, Google Drive, Dropbox (membre du GIFCT) et OneDrive (membre du GIFCT).

[39] Voir la section 3 du présent rapport.

[40] Voir la section 4.1 des profils suivants : Facebook, YouTube, Facebook Messenger, Instagram, Reddit, Twitter, Quora, Pinterest, Vimeo, Ask.fm, Twitch, Tumblr, VK, Medium, Odnoklassniki, Smule, KaKaoTalk, DeviantArt, Meetup, Dropbox et Wordpress.com.

[41] Voir la section 4.2 des profils suivants : Facebook, YouTube, WhatsApp, Facebook Messenger, Instagram, TikTok, Reddit, Twitter, Quora, Pinterest, Vimeo, LINE, Ask.fm, Twitch, Tumblr, VK, Medium, Discord, KaKaoTalk, DeviantArt, Meetup, 4chan et Wordpress.com.

[42] Par exemple, certains services utilisent une formule de type « peut recevoir un avertissement » (voir la section 4.1 du profil d'Ask.fm), qui laisse à penser que la notification n'est pas systématique et conduit à une part d'incertitude. Pinterest suit une approche similaire, indiquant que les utilisateurs sont avisés « dans la plupart des cas » (voir la section 4.1 du profil de Pinterest). Twitch, pour sa part, avertit les utilisateurs « selon la nature de l'infraction » (voir la section 4.1 du profil de Twitch). Autres exemples : Smule, Odnoklassniki et Wordpress.com (voir la section 4.1 des profils de Smule, d'Odnoklassniki et de Wordpress.com).

[43] Voir les sections 4 et 5 des profils suivants : WhatsApp, iMessage/FaceTime, WeChat, Instagram, QQ, TikTok, Weibo, iQIYI, Douban, LinkedIn, Quora, Snapchat, Pinterest, IMO, Ask.fm, VK, Haokan, Odnoklassniki, Smule, Meetup, MySpace et OneDrive. Le recours à des formulations de type « pourrait examiner... » ou « se réserve le droit d'examiner... », en particulier, est monnaie courante.

[44] Voir les sections 4 et 5 des profils suivants : WeChat, QQ, Youku Tudou, QZone TikTok, Weibo, iQIYI, Douban, Baidu Tieba, YY Live, Xigua, Huoshan et Haokan.

[45] Voir la section 7 des profils cités à la note précédente.

[46] Les statistiques sont disponibles à l'adresse : https://www.oberlo.co.uk/blog/tiktok-statistics.

[47] En d'autres termes, de prévenir toute utilisation de l'internet à des fins terroristes et la propagation des contenus terroristes et extrémistes violents sur leurs plateformes.

[48] Le GIFCT n'a pas précisé le sens de ce terme.

[49] À l'heure actuelle, les États-Unis, le Royaume-Uni, la France, le Canada, la Nouvelle-Zélande, le Japon, la Direction exécutive du Comité contre le terrorisme des Nations Unies et la Commission européenne sont membres du comité consultatif (GIFCT, 2019[30]). Pour plus d'informations sur la structure de gouvernance du GIFCT, voir : https://gifct.org/about/.

[1] Pour davantage d'informations sur les contenus violents et odieux en Australie et les dispositifs de blocage des fournisseurs d'accès à l'internet, consultez les documents suivants :

- eSafety Blog on Range of Christchurch Tools & Powers: https://www.esafety.gov.au/about-us/blog/christchurch-shifted-online-world-its-axis
- eSafety AVM Fact Sheet: https://www.esafety.gov.au/sites/default/files/2020-03/eSafety-AVM-factsheet.pdf
- eSafety ISP blocking Fact Sheet: https://www.esafety.gov.au/sites/default/files/2020-03/eSafety-ISP-Blocking-factsheet.pdf
- eSafety press release on landmark ISP blocking protocol: https://www.esafety.gov.au/about-us/newsroom/blocking-viral-spread-terrorist-content-online

Notes sur les profils

[1] Ce profil concerne la plateforme Facebook et non la société elle-même et ne porte donc pas sur Messenger, Instagram ou WhatsApp.

[2] Le programme YouTube Trusted Flagger a été créé par YouTube pour fournir des outils puissants aux utilisateurs, aux agences gouvernementales et aux organisations non gouvernementales (ONG) qui signalent avec efficacité des contenus qui enfreignent le Règlement de la communauté. https://support.google.com/youtube/answer/7554338?hl=fr

[3] Voir la section 3 du présent rapport.

[4] Il convient de préciser que les Conditions générales d'utilisation s'appliquent aux utilisateurs de QQ du monde entier, sauf s'ils appartiennent à l'une des catégories suivantes : (a) utilisateur QQ en Chine ; (b) ressortissant chinois utilisant QQ en quelque endroit que ce soit dans le monde ; (c) entreprise enregistrée en Chine utilisant QQ en quelque endroit que ce soit dans le monde. Ils doivent dans ce cas respecter les Conditions générales d'utilisation applicables en République populaire de Chine, disponibles à la page https://www.qq.com/contract.shtml.

[5] En dehors de la Chine, l'on accède à Qzone uniquement par l'intermédiaire de QQ International.

[6] Ces Conditions générales d'utilisation concernent les utilisateurs situés en dehors de la Chine. Les utilisateurs de QZone en Chine doivent respecter les Conditions générales d'utilisation applicables en République populaire de Chine, disponibles à la page https://www.qq.com/contract.shtml.