OECD publishing

CURRENT APPROACHES TO TERRORIST AND VIOLENT EXTREMIST CONTENT AMONG THE GLOBAL TOP 50 ONLINE CONTENT-SHARING SERVICES

OECD DIGITAL ECONOMY PAPERS
August 2020 **No. 296**

OECD
BETTER POLICIES FOR BETTER LIVES

Foreword

This report is the first in a series of two, to be issued one year apart. They are part of a larger project to develop a voluntary transparency reporting framework for terrorist and violent extremist content (TVEC) online. The two reports will take stock of the TVEC-related policies and procedures of the world's leading online platforms and other online content-sharing services, and how they have changed over the course of one year.

These reports are being written by Dr Tomas Llanos of University College London under the guidance of Jeremy West of the OECD. The author wishes to thank the delegates of the OECD Committee on Digital Economy Policy for their valuable feedback on earlier drafts, as well as the companies that reviewed their profiles to ensure accuracy. This report was approved and declassified by the Committee by written procedure on 5 May 2020.

The TVEC project is proceeding with the kind support of Australia, Canada, Korea and New Zealand.

Note to Delegations:

This document is also available on O.N.E under the reference code:

DSTI/CDEP(2019)15/FINAL

Table of contents

Executive Summary

Terrorists and violent extremists abuse the Internet to advance their agendas. They use apps, social media and other online content-sharing services to disseminate propaganda meant to glorify terrorism and violence, and to radicalise and recruit people. When terrorist and violent extremist content (TVEC) can be quickly replicated and distributed online at virtually no cost, terrorist and violent extremist ideologies spread more easily. This report is about the policies and procedures that the world's top 50 online content-sharing services have implemented with respect to TVEC, and it focuses on the issue of transparency.

The spread of TVEC online has contributed to numerous attacks, some of which have gone viral, bringing pressure on content-sharing services to do more to keep TVEC off of their services and stop it from spreading. However, without sufficient transparency and accountability, it may not only be difficult to understand how companies moderate TVEC online and how effectively their methods contain it, but the companies may also inadvertently curb fundamental rights such as the freedoms of expression, access to information and due process.

Thus, for example, the Christchurch Call (Christchurch Call, 2019[1]), which is a non-binding pledge to prevent the Internet from being used as a tool for terrorists and violent extremists, recognises that the measures online service providers take to counter TVEC online should be transparent, as should their community standards or terms of service. The Call also states that service providers should enforce those standards and terms in ways that are consistent with human rights and fundamental freedoms, and that they should "[implement regular and transparent public reporting, in a way that is measurable and supported by clear methodology, on the quantity and nature of terrorist and violent extremist content being detected and removed" (Christchurch Call, 2019[1]). The need to increase efforts to stop the spread of TVEC in a way that is transparent, accountable and compatible with fundamental rights and freedoms has also been recognised in the 2017 G20 Hamburg Leaders' Statement on Countering Terrorism (G20, 2017[2]), the 2019 G20 Osaka Leaders' Statement on Preventing Exploitation of the Internet for Terrorism and Violent Extremism (G20, 2019[3]), and the 2019 G7 Digital Ministers Chair's Summary (G7, 2019[4]).

This benchmarking report (the "Report") is part of the OECD's response to those calls for action. It summarises the current practices and procedures concerning TVEC of the global top 50 most popular online content-sharing services (the "Services"), identifying commonalities and trends in their approaches, and noting which ones issue transparency reports (TRs) on TVEC. The Report is an objective and factual snapshot, rather than a set of recommendations. It provides evidence to aid in understanding the Services' TVEC policies and procedures and determining the extent to which their implementation is transparent and accountable. The Report's findings can serve as a helpful baseline for discussing and building an effective cross-industry response to TVEC.

The Report also informs efforts led by the OECD, in collaboration with member countries, business, civil society and academia, to develop a multi-stakeholder, consensus-driven framework and set of metrics for voluntary transparency reporting by companies on TVEC online. The framework and metrics are intended to become part of a standardised template that all companies wishing to report on TVEC can use, and that all OECD members can accept.

The Report's key findings are:

- Only five of the top 50 online content-sharing services issue transparency reports specifically about TVEC.
- Those services that do publish TVEC transparency reports all do it idiosyncratically. They use different definitions of terrorism and violent extremism, report different types of information, use different measurement and estimation methods, and issue reports with varying frequency and on different timetables.

- The low number of reporting companies and the variation in what, when and how they report makes it impossible to get a clear and complete cross-industry perspective on the efficacy of companies' measures to combat TVEC online, as well as on the human rights impact that those measures have. Were more companies to issue TVEC transparency reports and include more comparable information, this situation could be improved.
- Thirteen of the top 50 online content-sharing services are Chinese and none of them issues TVEC transparency reports. That is not because there would be nothing interesting to put in them, though. Chinese regulations prohibit Internet content providers and publishers from displaying terrorist or extremist content. A 2017 Chinese cybersecurity law requires the transmission of banned content to be "immediately stopped" and obliges Internet companies to assist security agencies with investigations. The lack of transparency combined with the statutory requirements for handling TVEC may create tensions as Chinese services strive to expand in OECD countries.

Introduction

The rise of the Internet and the expansion of information and communication technologies have fundamentally transformed how individuals communicate, access and share information. This phenomenon has yielded numerous benefits while giving rise to new challenges and threats. Regrettably, the Internet and its enabling technologies are increasingly used to further terrorist ends (*terrorist use of the Internet,* TUI). While there is no universally accepted definition of terrorism or violent extremism, and, by extension, of terrorist and violent extremist content (TVEC), terrorists and violent extremists use apps, social media sites and other online content-sharing services to communicate and coordinate their actions across the globe and to disseminate TVEC,[1] including propaganda that is meant to glorify terrorism and violence and to radicalise and recruit individuals.[2] TVEC online is particularly concerning because digital information can be replicated and distributed at virtually no cost. TVEC online can therefore be widely circulated and even go viral within a short period of time, facilitating the spread of terrorist and violent extremist ideologies and propaganda. Unfortunately, the abuse of online content-sharing services has enabled terrorists and violent extremists to connect, grow, organise, and act with greater ease, speed, and breadth.

As a result of the proliferation of TVEC online, major companies such as Facebook, Twitter, and Google, as well as some smaller online content-sharing services, have come under public pressure to prevent terrorist and violent extremist groups from abusing their services. One of the main tools available to companies is known as "content moderation". Content moderation is generally understood as "the organised practice of screening user-generated content (UGC) posted to Internet sites, social media and other online outlets, in order to determine the appropriateness of the content for a given site, locality, or jurisdiction"[3] (Roberts, 2017[5]). When content moderation is employed, content found to violate a company's content or community standards or local legal frameworks can lead to actions such as content removal or blocking, suspension of the infringing account pending review, or a permanent ban from the platform. However, without transparency and appropriate accountability to their own terms of service and users, content moderation can curb fundamental rights and freedoms such as freedom of expression, access to information and due process.

On 15 March 2019, a gunman carried out and live streamed on Facebook a terrorist attack on two mosques in Christchurch, New Zealand. 51 people were killed, 50 injured, and the live stream was viewed around 4 000 times before being taken down (Christchurch Call, 2019[1]). The attack went viral and was subsequently found on the Internet despite the actions taken to remove it, exposing the need for greater efforts from both governments and tech companies to coordinate and implement collective actions aimed at the elimination of TVEC online (Christchurch Call, 2019[1]). Two months later, on 15 May 2019, a group of government leaders and major online service providers adopted a non-binding pledge, the Christchurch Call (2019[1]), to eradicate TVEC online and thereby prevent the Internet from being used as a tool for terrorists and violent extremists.

The Christchurch Call contains important commitments by a number of signatories, some of which are online service providers and some of which are governments. Online service providers committed to "[t]ake transparent, specific measures seeking to prevent the upload of terrorist and violent extremist content and to prevent its dissemination on social media and similar content-sharing services, including its immediate and permanent removal", "[p]rovide greater transparency in the setting of community standards or terms of service", "[e]nforce those community standards or terms of service in a manner consistent with human rights and fundamental freedoms", and "[i]mplement regular and transparent public reporting, in a way that is measurable and supported by clear methodology, on the quantity and nature of terrorist and violent extremist content being detected and removed." (Christchurch Call, 2019[1]) Governments, in turn, committed to work collectively with online service providers to develop "technical solutions to prevent the upload of and to detect and immediately remove terrorist and violent extremist content online", as well as to develop and implement "best practice[s] in preventing the dissemination of terrorist and violent extremist

content online", at all times respecting and protecting human rights that may be unduly impinged upon through business activities (Christchurch Call, 2019[1]). The need to increase efforts to stop the spread of TVEC in a way that is transparent, accountable and compatible with fundamental rights and freedoms has been echoed in other international fora, including the 2017 G20 Hamburg Leaders' Statement on Countering Terrorism (G20, 2017[2]), 2019 G20 Osaka Leaders' Statement on Preventing Exploitation of the Internet for Terrorism and Violent Extremism, which welcomed "online platforms' commitment to provide regular and transparent public reporting" (G20, 2019[3]), and the 2019 G7 Digital Ministers Chair's Summary (G7, 2019[4]).

This benchmarking report (the "Report") is part of the OECD's response to the international calls for action in those documents. The Report summarises the current practices and procedures concerning TVEC of each of the global top 50 most popular online content-sharing services (the "Services"). Based on this information, the Report identifies commonalities, developments and trends in the Services' approaches, placing emphasis on whether and to what extent the Services issue transparency reports (TRs) on TVEC. This Report aims to present an objective, neutral and factual snapshot of the Services' current approaches to TVEC. In doing so, it helps to provide an evidence base for understanding the Services' TVEC policies and procedures and determining the extent to which their implementation is transparent and accountable to the Services' own terms of service and to users. The Report's findings can serve as a baseline for discussing and building an effective cross-industry response to TVEC in line with international standards on human rights, including the right to freedom of expression, as suggested in the 2019 G7 Digital Ministers Chair's Summary (G7, 2019[4]).

The Report is also part of a larger OECD response to the aforementioned calls to action for countering TVEC online and reflects the Organisation's commitment to work together with business and other stakeholders. In particular, this Report informs efforts led by the OECD, in collaboration with member countries, business, civil society and academia, to develop a multi-stakeholder, consensus-driven framework and set of metrics for voluntary transparency reporting by companies on TVEC online. The framework and metrics will inform the development of a standardised template that all companies wishing to report on TVEC can use, and that all OECD members can accept.

It is important to note that the threat of TVEC online is moving to smaller platforms and apps, which may lack adequate human, technological and financial resources as well as the required expertise to address terrorist and violent extremist exploitation of their services and functionalities (Tech Against Terrorism, 2019[6]). Future benchmarking reports on TVEC and similar endeavours may therefore have to look beyond the global top 50 Services to enable better understanding of the situation of smaller services, and thus a fuller picture of the problem, to keep up with terrorist and violent extremist groups' strategies.

Section 1 explains the scope of the research contained in this Report and its methodology. Section 2 sets forth the current state of play amongst the global top 50 Services with regard to their policies and efforts on TVEC and identifies issues, similarities and trends. Section 3 provides an overview of the industry-led Global Internet Forum to Counter Terrorism (GIFCT), an initiative by a group of technology companies included in this Report to combat terrorist and violent extremist exploitation of their platforms. Section 4 outlines current laws and regulations concerning TVEC, and other related legal or regulatory proposals, on a worldwide level.

1 Scope, Methodology and Research Design

This Report explores the existing policies, procedures and practices relevant to TVEC of 50 social media platforms, online communications services, file sharing platforms, and other online services whose businesses enable the uploading, posting, sharing and/or transfer of digital content and/or facilitate voice, video, messaging or other types of online communications. To determine which entities to include in the research, it made sense to include the most widely used or "popular" Services. To do this, attempts were made to measure a Services' popularity based on a common metric. Unfortunately a common metric proved elusive because of the diversity of products and services offered, and the distinct purposes and audiences served[4]. For example, the popularity of social media platforms can be measured based on the monthly active users (MAU) metric. However, that metric is unavailable for file sharing services and online encyclopaedias, the popularity of which can be established based on different measures.

To address this, Services were divided into three categories:

a. social media, video streaming services and online communications services;
b. cloud-based file sharing services; and,
c. an "other" category, which includes a content management service and an online encyclopaedia.

Within each category, the most "popular" Services were chosen. To determine popularity in each category, the following metrics were employed:

- Social media platforms, video streaming services and online communications services were chosen based on monthly average users (MAU). The MAU metric is commonly used by industry analysts and investors to determine a service's popularity and growth,[5] and constitutes a reliable measure to rank with a fair degree of precision the relative size of services that thrive on user engagement.
- Cloud-based file sharing services were chosen based on indicative market shares, a metric that is frequently used to determine the relevance of firms in a given industry segment.
- The third part includes two important services whose popularity cannot be determined relative to the other two groups; however, their undoubted relevance warranted their inclusion. Their importance was determined on the basis of data (indicative market share and monthly pageviews) that reveal their reach and/or usage.

A list of the 50 Services included in this benchmarking was assembled (Annex A).

The research proceeded in three main steps. First, a standardised profile template was devised, addressing all the fields of information that comprise the scope of the research. One profile per Service was developed based on each Service's publicly available terms of service (ToS), community guidelines and policies, blogs, service agreements and other official information (governing documents).[6] The Services were contacted and given an adequate period of time to provide their feedback on the accuracy of the profiles, as well as any relevant additional information.

Second, the profiles were updated based on the Services' responses. The final versions of the profiles appear in Annex B.

Third, commonalities, developments and trends in the Services' approaches to TVEC were identified. These findings are presented in Section 2 of this Report.

The Report focuses on collecting information on the policies and practices of the Services in several important areas:

a. definition of terms like terrorist/terrorism and violent extremist/violent extremism;
b. identification and removal of TVEC, including policies on enforcing compliance with terms and conditions of service, on removals, on sanctions, and whether there are appeals processes;
c. consequences for user breaches of terms of service/community guidelines and standards;
d. voluntary issuance of transparency reports (TRs) concerning TVEC including their content, methodology and frequency.

The Report will be updated in 12 months, repeating the three steps outlined above, in order to observe changes to the Services' policies and procedures concerning TVEC, including the issuance of TRs on TVEC.

2 Commonalities, Developments and Trends in the Services' Approach to TVEC

Vague Descriptions of TVEC and Related Concepts, and Diverging Approaches to Identifying 'Terrorist Organisations'

There is no universally accepted definition of terrorism or violent extremism, and, by extension, of TVEC. Accordingly, it is unsurprising that the Services do not use the term TVEC, nor do they define terrorism and violent extremism in uniform ways. However, whilst the majority of the Services explicitly ban content that, to a greater or lesser degree, can be considered TVEC or TUI,[7] only five Services attempt to define terrorism, violent extremism and related concepts with sufficient detail to understand the scope of such terms, providing examples where appropriate.[8]

Different degrees of specificity can be seen in the remaining 45 Services' approaches to describing terrorist content and/or violent extremist content. Nineteen Services explicitly ban the use of their technologies to foster terrorist aims, using the terms terrorist/terrorism, violent extremists/violent extremism and similar expressions.[9] Fifteen Services conflate hate speech and/or violent or graphic content with TVEC.[10] Sixteen Services use broad and/or vague descriptions of prohibited conduct, which descriptions can be interpreted as supersets encompassing TVEC.[11]

The Services also have different approaches to identifying what a terrorist organisation is, which has repercussions on what is deemed TVEC. For example, Facebook states that it enforces its Community Guidelines as applied to terrorist activities and groups both regionally and globally. Prior to November 2019, its transparency reports measured only the actions Facebook took on terrorist propaganda related to ISIS, al-Qaeda and their affiliate groups.[12] At that point, Facebook expanded its reporting metrics to include the company's efforts against all terrorist organisations (Facebook, 2020[7]). Other firms such as Twitter and Automattic (WordPress.com's parent company) do not mention which particular groups their efforts to counter TVEC focus on. Some Services rely on lists of terrorist organisations issued by governments. YouTube, for example, specifies that content that violates their policies against violent extremism includes material produced by 'government-listed' foreign terrorist organisations,[13] without indicating what specific government(s) it is referring to.[14] Microsoft, in turn, indicates that it considers terrorist content to be material posted by or in support of organisations included on the Consolidated United Nations Security Council Sanctions List.[15]

Greater definitional precision and more detailed delineations and explanatory sections on what the Services consider TVEC are essential for determining whether the Services are addressing the same, similar or different content, particularly in light of the absence of a widely accepted definition of TVEC. Also, without clearer indications of what groups the Services' reporting and approaches focus on, it is difficult to understand whom their TVEC moderation efforts are affecting and how. Research has shown that there are profound differences in how companies approach different violent extremist groups and that not all platforms target the same groups equally.[16]

Transparency Reports[17] Expressly Addressing TVEC Are Uncommon

The practice of reporting information on how companies moderate and remove content based on their own ToS and policies generally, and based on their anti-terrorism and anti-violence policies in particular, is hardly widespread. Of the 23 Services profiled in this Report that issue any transparency reports at all,[18] only five (Facebook, YouTube, Instagram, Twitter and Automattic) issue reports specifically about TVEC.[19]

It is axiomatic that reliable, comprehensive information is essential for understanding a problem and the progress being made towards a solution. Such information is also necessary to inform productive deliberation and debate. Transparency reporting has emerged in different areas[20] to serve these and related ends. Online content moderation is a case in point, albeit a nascent one. The Internet has opened multiple avenues for individuals to express their opinions, views and beliefs and access information about myriad topics. At the same time, the online platforms and services that enable and carry individuals' expressions and information can be *de facto* speech gatekeepers, empowered and sometimes legally bound to remove or block users' content. Transparency reporting on content removal by Internet companies increases visibility into this role, allowing insights into whether these firms' practices are achieving (or at least are not inconsistent with) policy goals (set by both the companies themselves and governments), including whether they are respecting users' fundamental rights and freedoms, such as freedom of expression (New America, n.d.[8]).

Google became the first Internet firm to publish a TR in 2010, focusing on U.S. and non-U.S. government requests for content takedowns (Google, 2010[9]). 'Behind the curtains' content removals at the request of governments are detrimental to freedom of expression and prevent the free flow of information. Reporting of this type was intended to deter government censorship and uphold fundamental rights and freedoms (Google, 2010[9]). Other companies including LinkedIn, Microsoft and Twitter followed suit over the next several years, but transparency reporting became a more common practice in 2013 (New America, n.d.[10]). Government surveillance revelations called into question the handling of private data by US Internet firms, causing a consumer crisis to which major firms responded by publishing detailed reports about government demands for data (New America, n.d.[10]).

Since then, many technology companies have published TRs to document the scope and extent of government requests for user data. Over time, transparency reporting by some companies expanded to include information on topics including intellectual property-related takedowns, government and legal requests for content moderation and removal, and child exploitation. At this time, however, only a handful of Services issue transparency reports that specifically discuss their policies and practices for moderating terrorist and/or violent extremist content.

Differences between Current TVEC Transparency Reports

The definitions used and the kinds of information included in the five TVEC TRs currently issued, as well as their timing and frequency, are all different from one another. Consider, for example, the types of information being reported. The TRs issued by Automattic, a content management service (e.g. blog hosting), disclose notices of terrorist content it receives from government Internet Referral Units (IRUs). The TRs include the total number of notices received, the total number of notices that resulted in suspended sites, and the total percentage of notices that resulted in suspended sites. The TRs also provide monthly breakdowns for each of these categories.[21]

Twitter reports the number of requests received from governments worldwide to remove content in violation of Twitter's policies, including its anti-terrorism policy. Twitter also discloses the number of accounts reported for possible violations of its rules; the number of accounts on which it took action based on six categories of violations; the number of accounts suspended for violations related to the promotion of terrorism, and the percentage of these violations detected by Twitter's internal tools.[22]

YouTube discloses the number of content removal requests by governments based on six categories; the number of channels removed, sorted by the basis for removal; the number of videos removed by source of first detection; the percentage of videos first flagged through automated methods, with and without views; the percentage of videos removed, sorted by the basis for removal (including YouTube's violent extremism and hate speech policies); the number of comments removed, and the percentage of comments removed sorted by the source of first detection (automated flagging or human flagging).[23]

Facebook reports how prevalent terrorist propaganda violations on Facebook were; how much content it took action on; the percentage of the violating content it actioned before users reported it; the number of appeals against the decisions to take an action on specific content; and the amount of content it restored after removing it.[24] Instagram reports the first three metrics.[25]

These reports are, of course, useful. They provide much-needed insights into some of the most popular Services' TVEC moderation and removal efforts, perspectives on how much TVEC is showing up and how much content is mistakenly blocked or taken down, among other things. However, the significant variance among the Services' reporting complicates comparison and analysis.

For example, whilst Twitter and Automattic disclose the number of government notices they receive for terrorist-related content on their services, Facebook, Instagram and YouTube do not provide comparable information. Consistent reporting of such notices could facilitate the assessment of countries' efforts to enforce their laws and policies to counter TVEC online.

Moreover, whilst YouTube reports the number of videos removed by source of first detection (for example, automated flagging or human flaggers), this number is not broken down by category of policy violation (for example, nudity as opposed to terrorist propaganda).[26] Accordingly, it is not possible to determine what percentage or volume of terrorist propaganda content YouTube removed based on automated detections versus human reporting. YouTube does report the number of human flags by reason for removal, including the promotion of terrorism;[27] however, it does not disclose how much of that content was subsequently removed. This limits analysis of YouTube's approaches to terrorist propaganda.

Similarly, Facebook and Instagram, though they have made very substantial efforts with their transparency reporting, disclose the volume of content they identified and actioned before users reported it,[28] but they do not specify whether that content was identified by automated tools, human reviewers or in some other manner. For its part, Twitter does indicate the percentage of accounts it suspends based on violations of their anti-terrorism policy detected by their proprietary tools,[29] thereby providing a useful indication of such tools' significance in its counter-terrorism efforts.

Facebook's approach to reporting TVEC moderation is the most comprehensive featuring clearly defined metrics, helpful descriptions of calculation methodologies, and explanations as to why the metrics are important for understanding moderation of different types of content.[30] It is important to understand that Facebook's prevalence metric for terrorism-related content is an estimate based on samples of content across different areas of the platform, such as Groups and News Feeds. Although there are good reasons for that, because it is an estimate, Facebook's prevalence metric may be inaccurate. Indeed, Facebook has been accused of understating the prevalence of terrorist content on its platform (Engineering & Technology, 2019[11]). Since content cannot have much of an impact unless users view it, an accurate prevalence metric is of the essence to determine the impact of a Service's efforts to counter TVEC. However, no other Service in the top 50 reports a TVEC-related prevalence metric at all. [31]

There would be significant challenges to achieving absolute uniformity in voluntary transparency reporting on TVEC. The content of a TR on TVEC depends to a certain extent on the Services' content moderation targets and priorities, which in turn may depend on the content that is typically shared on their platforms. This could be why, for example, Twitter looks at the number of accounts, whereas Facebook looks at the number of pieces of content. YouTube looks at both comments and accounts.[32] In addition, the companies included in the top 50 list offer different services and operate in different ways. Depending on the business

model, access to certain information by the Service operator, and as a consequence the calculation of metrics that are of great relevance in other Services' TRs on TVEC, may be difficult. For example, an electronics communications app with two-way encryption, such as Telegram, cannot readily access the content shared amongst its users and consequently would have a hard time calculating a prevalence metric such as that reported by Facebook. Nevertheless, aspiring to and working towards a higher level of standardisation in TRs on TVEC, to the greatest extent practicable, could enable better identification and assessment of impact and best practices in the Services' TVEC moderation and removal efforts.[33]

Staff Member Moderators, User-Moderators and Automated Tools

Staff member moderators, user-moderators and automated tools may all be used to detect and remove objectionable content, including TVEC. Each approach has strengths and weaknesses, so choosing one or a combination of them entails a trade-off that the Services must make in consideration of the type of service they offer, the size of their user base, their technological prowess, their financial resources, and other factors.

Staff member moderators (including contractors) are individuals hired to monitor and moderate content on their employer's platforms and services. These human moderators tend to be more costly than the other two approaches and are slower than automated tools. As a result, Services with large user bases such as Facebook and YouTube use staff member moderators in combination with the other approaches, since they cannot effectively and efficiently monitor all the content shared on those platforms. Staff member moderators are, however, able to make nuanced decisions, which is particularly important when policing content like TVEC that does not have a clear or simple definition.

Automated tools, on the other hand, are the opposite in many ways. They are faster and cheaper than staff member moderators (at least in terms of marginal costs, though the fixed costs may be high), so they can deal more efficiently with large volumes of content. Automated tools are not very good at taking subtle contexts into account, though, and they reflect the biases of their designers (OFCOM, 2019[12]). Therefore, human moderators are still needed to take nuances into consideration and correct any biases.

User moderation is a virtually costless method in which content moderation is outsourced to a volunteer corps of users (Crawford, 2014[13]). This approach is sometimes criticized because users bring their own biases and interpretations of community guidelines to their decisions about who and what to report. Moreover, user moderation is particularly susceptible to abuse.[34]

The majority of the Services rely on staff member moderators to detect violations of their ToS and policies.[35] Ten of the fifty Services analysed employ systems that rely on users as moderators.[36] At least[37] twenty-one employ automated tools to detect violations of their ToS and policies.[38]

As will be seen in Section 3, the joint tech innovation efforts channelled by the GIFCT have resulted in the creation of a shared industry database of 'hashes' or unique digital 'fingerprints' for TVEC that GIFCT members have removed from their online services. Broadly speaking, the hash of TVEC identified and removed by one user of this database (i.e. a 'Hash Sharing Consortium' member) is shared with the other members to enable them to automate the identification and moderation of that TVEC on their own platforms, and even to block that TVEC before it is posted. Adoption of automated tools using the GIFCT's hash database is likely to increase in the future, as more firms are joining the GIFCT and its Hash Sharing Consortium. Amazon, Dropbox, LinkedIn, Pinterest and WhatsApp recently became GIFCT members (GIFCT, 2019[14]). All of the GIFCT's members are included in this research. In addition, of the fifty Services listed in Annex A, nine participate in the GIFCT's Hash Sharing Consortium.[39]

Notification, Enforcement and Appeal Mechanisms and Processes

Twenty-one Services have mechanisms for notifying users in case of potential violations of their ToS and other governing documents.[40] Twenty-three Services have appeal processes in place in respect of content moderation decisions and other measures applied under their governing documents.[41] Services take different approaches in notifying users of enforcement decisions taken against them or their accounts.[42]

The remaining Services either have no appeal processes or do not provide public information in this regard. That may generate suspicions of moderation decisions that cause the over-removal of content without sufficient notice, in violation of the Services' ToS.

In the case of twenty-two Services, a clear understanding of whether they review content proactively and/or reactively to determine compliance with their ToS and policies is difficult to obtain.[43] Some Services may be reluctant to acknowledge monitoring activities currently in place, or alternatively, remain vague to deflect criticism if they do not monitor content at all.

Disclosure by Chinese Platforms

The Chinese Services generally provide limited information with respect to their content moderation practices and processes for enforcing their ToS and policies.[44] With the exception of TikTok, none of them issues TRs of any kind.[45]

This tendency among Chinese Services may be explained by the regulatory framework, which prohibits Internet content providers and Internet publishers from posting or displaying content that, among other things, violates the laws and regulations of the People's Republic of China (hereinafter, "China"), impairs the national dignity of China, or contains terrorist or extremist content (Baidu, Inc., 2017[15]). It should be noted here that "extremist" content is broader than *violent* extremist content. The regulatory environment in China creates a system of intermediary liability under which online content-sharing services have legal responsibility for content control (Knockel, 2018[16]). Moreover, the Chinese government has introduced successively stricter requirements in an effort to increase its control over Internet traffic and content. A new cybersecurity law came into effect in June 2017, increasing censorship requirements (for example, the transmission of banned content must be "immediately stopped"), mandating data localisation, codifying real-name registration requirements for Internet companies, and obliging them to assist security agencies with investigations (Creemers, 2018[17]). Companies are expected to invest in staff and filtering technologies to moderate content and stay in compliance with governmental rules (Knockel, 2018[16]). Failure to comply with these requirements may result in the revocation of licenses to provide Internet content and other services, fines and/or the closure of the concerned services.

To manage increasing government pressures, Chinese Services have been investing more in both filtering technologies and human resources to moderate content. Global Times, a Chinese state media outlet, reported that tech companies are expanding their human censor teams and developing artificial intelligence tools to review "trillions of posts, voice messages, photos and videos every day" to make sure their content is in line with laws and regulations (Zhang, 2018[18]). Media reports indicate that the majority of Chinese platforms are equipped with a keyword filter that allows them to automatically censor sensitive information before it is published. State censorship authorities constantly update a list of keywords and distribute it to platform operators (Wang, 2019[19]).

Against this background, Chinese Services' limited disclosure regarding content moderation and monitoring seems to align with the domestic regulatory framework. If they publicly acknowledge that they closely monitor users' activities and remove any content that violates applicable laws and regulations, they may make their Services less attractive on privacy and freedom of speech grounds. Also, an acknowledgment of this type would highlight the absence or at least vagueness of the published rationales

for those removals. However, to comply with such laws and regulations, the Services are bound to moderate content in close cooperation with the government. WeChat provides a good example of these conflicting interests. It has publicly stated that it does not interfere with or analyse the content of user chats (Corfield, 2018[20]); however, research has shown that content and messages are routinely censored on WeChat (Ruan, 2016[21]).

This tension could be problematic in the context of Chinese Services' international expansion ambitions, as non-Chinese audiences may be particularly suspicious of Chinese government-driven surveillance and censorship practices (Washington Post, 2019[22]). The example of TikTok shows that penetration of international markets requires alignment with Western transparency standards on content moderation, policies, and practices. With a growing user base in the United States and other OECD countries,[46] TikTok has been at pains to ensure that its content moderation practices are not based on 'sensitivities to China', asserting that TikTok and its Chinese version, Douyin, are not conflated with one another (Wired, 2019[23]). In October 2019, TikTok announced it was summoning external experts to review some of its content moderation policies (TikTok, 2019[24]), and in December 2019 it released its first TR, disclosing 298 legal requests for user information and 26 government content removal requests during the first half of 2019 (TikTok, 2019[25]). The TR does not cover Douyin, however. It seems that the tension noted above is leading to a different parallel treatment for Tiktok's domestic and international versions: a more 'open' and transparent approach for TikTok, eschewing ties to Chinese regulatory requirements, and a traditional, more secretive approach for Douyin in line with China's regulatory framework. A similar dual system has been observed between WeChat and its domestic version Weixin (Ruan, 2016[21]).

3 GIFCT

Partly as a response to mounting pressure from governments and the public to curb the online propagation of TVEC, Facebook, Microsoft, Twitter, and YouTube formed the Global Internet Forum to Counter Terrorism (GIFCT) in July 2017. Amazon, Dropbox, LinkedIn, Pinterest and WhatsApp later became members (GIFCT, 2019[14]).

GIFCT's mission statement is to 'prevent terrorists and violent extremists from exploiting digital platforms.'[47] Its goals are to

- improve the capacity of a broad range of technology companies, independently and collectively, to prevent and respond to abuse of their digital platforms by terrorists and violent extremists
- enable multi-stakeholder engagement around terrorist and violent extremist misuse of the Internet and encourage stakeholders to meet key commitments consistent with the GIFCT mission
- encourage those dedicated to online civil dialogue and empower efforts to direct positive alternatives to the messages of terrorists and violent extremists, and
- advance broad understanding of terrorist and violent extremist operations and their evolution, including the intersection of online and offline activities (GIFCT, 2017[26]).

To achieve its goals, the GIFCT employs four, inter-related strategies: joint tech innovation, knowledge-sharing, conducting funding and research, and content incident protocol. Information-sharing efforts fall under the GIFCT's joint tech innovation strategy, focusing on building shared technology for use within the tech industry to prevent and disrupt the spread of terrorist content online. These efforts have resulted in the creation of a shared industry database of 'hashes' — unique digital "fingerprints" — of known terrorist images and videos. The image or video is "hashed" in its raw form and is not linked to any source platform or user data. Hashes appear as a numerical representation of the original content and cannot be reverse engineered to create the image and/or video. A platform has to find a match with a given hash on their site in order to see what the hash corresponds with. It is up to each company using this hash database to determine how they use the database, depending on their own terms of service, how their platform operates, and how they employ technical and human capacities. (GIFCT, 2019[27]). GIFCT claims that this collaboration is resulting in increased efficiency in the enforcement of its member's counterterrorism policies (GIFCT, 2017[28]). The database currently contains more than 200 000 hashes, which member companies can use to identify and remove matching content that violate their respective policies or, in some cases, block terrorist content before it is even posted (GIFCT, 2017[28]).

Companies that use the hash database comprise the 'Hash Sharing Consortium'. Its current members are Microsoft, Facebook, Twitter, YouTube, Ask.fm, Cloudinary, Instagram, JustPaste.it, LinkedIn, Verizon Media, Reddit, Snap, and Yellow (GIFCT, 2019[27]).

GIFCT released its first transparency report in July 2019, clarifying certain aspects of its operations and cross-sector progress and relations (GIFCT, 2019[27]). GIFCT disclosed that, for the purposes of the hash sharing database, and to find an agreed upon common ground, founding companies in 2017 decided to define terrorist content based on content relating to organisations on the United Nations Security Council

Consolidated List. They also agreed upon a basic taxonomy of the content posted that relates to these listed organisations. The taxonomy includes the following labels that are applied to the content when a company adds hashes to the shared database:

- Imminent Credible Threat: A public posting of a specific, imminent, credible threat of violence toward non-combatants and/or civilian infrastructure.
- Graphic Violence Against Defenceless People: The murder, execution, rape, torture, or infliction of serious bodily harm on defenceless people (prisoner exploitation, obvious non-combatants being targeted).
- Glorification of Terrorist Acts: Content that glorifies, praises, condones or celebrates attacks after the fact.
- Recruitment and Instruction: Materials that seek to recruit followers, give guidance or instruct them operationally.
- New Zealand Perpetrator Content: Due to the virality and cross-platform spread of the Christchurch attacker's manifesto and attack video, and because New Zealand authorities deemed all manifesto and attack video content illegal, the GIFCT created a 'crisis bank'[48] to mitigate the spread of this content.

GIFCT also provided information on its URL-sharing initiative. Companies only have jurisdiction to remove the primary source content from what is hosted on their services, meaning they can remove a post, but the source link and hosted content remains intact on a third party platform. In 2018, Twitter began a program to share URLs to the platforms that were linked to Twitter posts associated with terrorist content. GIFCT expanded this program starting in January 2019 to allow GIFCT companies to safely share suspicious URLs with the industry partner to which the URL belongs. The one-to-one sharing allows the notified platform to review the link in accordance with its own terms of service so it can decide whether the content violates them (GIFCT, 2019[27]).

In a meeting held at the United Nations on 24 September 2019, tech companies and world leaders announced a number of measures to implement the Christchurch Call to Action, including an overhaul of the GIFCT to make it an independent body that will drive much of the tech sector's work on implementing the Call. The GIFCT is now re-established as an independent non-profit 501(c)(3) organisation in the United States (GIFCT, 2019[29]). Other reforms are in progress, including the recruitment of an independent Executive Director to lead GIFCT and be responsible for coordinating all operations, including core management, program implementation and fundraising. The GIFCT will be governed by an industry-led Operating Board, which will work closely with a broad multi-stakeholder Forum and an Independent Advisory Committee (GIFCT, 2019[30]). The Committee will be chaired by a non-governmental representative and will include members from civil society, government and inter-governmental entities.[49]

The relaunch of the GIFCT included a revision of its mission mandate. The GIFCT is now concerned with both terrorist *and* violent extremist content online, and its planned endeavours include

- investing in new technology
- promoting alternative narratives and positive interventions
- being more inclusive and transparent, with multi-stakeholder engagement across its activities, bringing civil society to the heart of the fight against TVEC, and
- establishing working groups that will focus on six areas:
  - transparency
  - crisis response
  - legal frameworks
  - technical approaches

- algorithmic outcomes
- academic and practical research

In addition, the GIFCT is providing support for the creation of the Global Network on Extremism and Technology (GNET), which will bring together an international consortium of leading academic institutions and experts with core institutional partnerships from the United States, the United Kingdom, Australia, Germany and Singapore to study and share findings on combating terrorist and violent extremist use of digital platforms.

4 Laws and Regulations on TVEC Online that Have Been Enacted or Are Currently under Consideration

Because terrorist and violent extremist groups misuse online services to disseminate propaganda and recruitment material, technology companies have faced increased pressure from governments and institutions around the world to ramp up efforts to combat the groups' operations. Concerned that, to date, industry efforts to counter TVEC have been inadequate, some governments have begun to propose and enact laws and regulations, and to implement other initiatives, to curb the online propagation of TVEC. This Section provides an overview of such responses, and also summarises certain statutes, laws and regulations that are of great relevance to the fight against TVEC and TUI.

Australia

In the aftermath of the Christchurch terrorist attacks, the Australian Parliament responded by passing the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (Act), which came into force on 6 April 2019 (Australian Government, 2019[31]). The Act adds new offences to the Criminal Code concerning online abhorrent violent content.

Abhorrent violent material is audio, visual, or audio-visual content that records or streams abhorrent violent conduct, produced by the perpetrator(s) of that conduct (or an accomplice) that a reasonable person would consider offensive in the circumstances. Abhorrent violent conduct is defined to mean murder or attempted murder, a terrorist act, torture, rape or kidnapping. There is no requirement that the person needs to be convicted of an offence in order for their conduct to constitute abhorrent violent conduct. For the purposes of the Act, it is immaterial whether or not the abhorrent violent material has been altered (for example, through the superimposition of other material). However, if the material is altered to such an extent that it no longer meets the criteria of abhorrent violent material (through appropriate editing), it will not be captured by the legislation.

Under the Act, it is an offence for an Internet service provider, content service or hosting service to fail to refer to the Australian Federal Police (AFP) 'within a reasonable time' abhorrent violent material that the provider is aware could be accessed through or on their service, where the underlying conduct occurred or is occurring in Australia. The term 'reasonable time' is not defined in the Act. However, the Explanatory Memorandum states that this will ultimately be a question for the trier of fact (for example, a jury) and will depend on factors such as the volume of the material (for example, how frequently it was posted and re-posted) and the capacity and resources of the service provider (that is, its technical removal capabilities).

In addition, under the Act it is an offence for a content or hosting service provider to fail to expeditiously remove from their content or hosting service abhorrent violent material that is reasonably capable of being accessed in Australia (regardless of where the service itself is located). The question of whether or not

specific content has been 'expeditiously removed' is, again, a matter for the trier of fact and will depend on factors such as the type and volume of the material and capabilities and resources of the service provider.

The Act also empowers the eSafety Commissioner to issue notices to content or hosting service providers to notify them that their services could be used at the time of issuing the notice to access abhorrent violent material.

Moreover, the Australian Taskforce to Combat Terrorist and Extreme Violent Material Online (the Taskforce) was established in March 2019, the objective of which is to provide advice to Government on practical, tangible and effective measures and commitments to combat the upload and dissemination of terrorist and extreme violent material (Department of the Prime Minister and Cabinet, 2019[32]). Fulfilling its remit, the Taskforce issued a report on 30 June 2019, identifying actions and recommendations that fall into one of five streams: prevention; detection and removal; transparency; deterrence; and capacity building. Some of such actions and recommendations include:

a. Digital platforms must continue to develop and report to the Australian Government on the ongoing development of technical solutions that seek to prevent terrorist and extreme violent material from being uploaded onto their services,

b. Digital platforms must work with other members of the GIFCT to strengthen the hash-sharing database and the URL-sharing consortium, with an aim to align, to the extent possible, with the categories of violent content prohibited by platforms under their respective community standards and terms of service, such as graphic violence, violent content or gore.

c. Digital platforms must have in place clear, efficient appeals mechanisms that provide users with the ability to challenge moderation decisions regarding terrorist and extreme violent material.

d. Overseen and managed by the Australia-New Zealand Counter-Terrorism Committee, digital platforms and relevant Australian Government agencies should convene a 'testing event' in 2019-20, simulating a scenario which will allow all parties to gauge whether industry tools, and Government processes, are working as intended, particularly as they mature in response to technology and increased investment in content moderation.

e. The Australian Government should pursue legislative amendments to establish a content blocking framework for terrorist and extreme violent material online in crisis events.

f. Digital platforms should publish reports (at least half yearly) outlining their efforts to detect and remove terrorist and extreme violent material on their services. These reports are intended to demonstrate the nature and extent of actions being taken by platforms, and could include:

   - the number of items flagged by users for potential violations of policies against the promotion of terrorism or extreme violent content;
   - the total number of items removed by the digital platform
   - the number and entity type (e.g. video, channel) of items of terrorist content and extreme violent content removed by the platform;
   - examples of content flagged for promotion of terrorism or extreme violence that did and did not violate the platform's guidelines;

- the number of items of terrorist content and extreme violent content that were flagged or identified by the platforms' systems;
- the total number of items of terrorist content and extreme violent content that were subject to moderation, broken down by those that were flagged by users, systems, other sources, and the total volume of content removed; and
- the average time taken to review and action flagged items of terrorist content and extreme violent content, or the number of times flagged terrorist content or extreme violent content was viewed by users before action was taken.
- the implementation of appropriate checks on live-streaming aimed at reducing the risk of users disseminating terrorist and extreme violent material online (Department of the Prime Minister and Cabinet, 2019[32]).[50]

European Union

Measures in the European Union to tackle illegal content online have evolved over time, moving from voluntary initiatives through to the current negotiations for binding measures in the form of a proposal for a regulation on preventing the dissemination of terrorist content online.

The EU Internet Forum was launched in December 2015 by the European Commission (hereinafter, "EC"), with the aim of addressing the misuse of the Internet by terrorist groups. The Forum is a voluntary partnership of EU Home Affairs Ministers, Internet industry representatives and other stakeholders.

In May 2016 the EU *Code of Conduct on Countering Illegal Hate Speech Online* was launched, with Facebook, Twitter, YouTube and Microsoft agreeing to adhere to the Code. Snapchat, Instagram, Dailymotion, Google+ and Jeuxvideo subsequently agreed to adhere to the Code. The Code aims to ensure requests to remove content are dealt with promptly (European Commission, 2019[33]).

The EC *Communication on tackling illegal content online* (European Commission, 2017[34]) was delivered in September 2017. The Communication provided guidance on the responsibility of online service providers with respect to all types of illegal online content, as defined by national and EU law.

In March 2018 the EC *Recommendation on measures to effectively tackle illegal content online* (European Commission, 2018[35]) translated the political commitment of the Communication into a non-binding legal form. The Recommendation included proposals for stronger procedures for more efficient removal of illegal content, and increased protection against terrorist content online.

The September 2018 EC Proposal for a Regulation on preventing the dissemination of terrorist content online contained a number of proposed requirements, including for:

a. platforms to take down terrorism-related content within one hour of receiving a removal order
b. hosting service providers to take proactive measures to remove terrorist material from their services, including by deploying automated detection tools; and
c. proposed penalties for platforms who fail to meet these requirements of up to 4 per cent of their global revenue.

In addition, the proposal imposed on hosting service providers the obligation to publish annual transparency reports on actions taken against the dissemination of terrorist content, and identified four aspects that TVEC transparency reports should contain at a minimum:

a. information about the hosting service provider's measures in relation to the detection, identification and removal of terrorist content;

b. information about the hosting service provider's measures to prevent the re-upload of content which has previously been removed or to which access has been disabled because it is considered to be terrorist content;
c. number of pieces of terrorist content removed or to which access has been disabled, following removal orders, referrals, or proactive measures, respectively; and
d. overview and outcome of complaint procedures.

The revised Audio-visual Media Services Directive was adopted in November 2018 and included, amongst other things, obligations for video-sharing platforms to enable flagging of terrorist content uploaded by their users.

The Council adopted a general approach on 6 December 2018. The European Parliament voted on its first reading of the Proposal for a Regulation on preventing the dissemination of terrorist content online on 17 April 2019, with extensive amendments to the Commission proposal. For example, it narrowed the definition of terrorist content, rejected some provisions such as the use of proactive measures and mandated responses to Internet Referral Units (IRUs), and reduced the scope of the regulation to public content posted online (as the definition of "hosting service provider" no longer applies to cloud and infrastructure services and electronic communication services) (European Parliament, 2019[36]). The Council of the European Union and the European Parliament, as well as the EC, are currently in trilogue negotiations working towards agreement on the final text, which is likely to amend the details and extent of the provisions mentioned above.

France

France has taken several steps to counter TVEC online, following high profile domestic terrorist attacks; including the Charlie Hebdo terror attack in January 2015 and the November 2015 Paris attacks. Measures under the *Law on Confidence in the Digital Economy 2004* and other counterterrorism laws include:

a. creating a blacklist of sites containing material that incites or condones terrorism;
b. requesting the hosts of such content to remove it; and
c. requesting that ISPs block websites containing infringing material (if the material not been removed in the 24-hour period following a removal request).

Delisting online content from search results is another method used to counter the spread of pro-terrorist content (Freedom House, 2018[37]).

In March 2019, the French Government introduced a Bill to parliament to tackle cyber-hate (Government of France, 2018[38]). The Bill was amended several times, and included provisions that would:

a. Require high traffic platform operators to remove manifestly illicit hate material, including incitement to hate or violence and racist or religious bigotry, within 24 hours of notification or risk a fine up to EUR 1.25 million
b. Require the platforms to remove terrorism and paedophilia-related content within one-hour after receiving instructions from government authorities to do so, or be subject to fines of up to €1.25m or up to 4% of social networks' and other online content providers' global revenue
c. Require a "single reporting button" common amongst platforms, enabling users to flag abuse
d. Require platforms to have adequate human and technology resources to meet obligations
e. Provide clear information on available remedies to victims of cyber-hate

f. Require platforms to designated a legal representative in France to assist and liaise with authorities, and
g. Introduce new powers to empower the administrative authority/regulator.

The Senate passed the Bill in March 2020 and the National Assembly approved it in May 2020. However, on 18 June 2020, the Constitutional Council found the core of the law to be unconstitutional because certain obligations it imposed on Internet platform operators infringe on the freedom of expression and communication (Conseil Constitutionnel, 2020[39]).

Germany

Regulating online hate speech and other forms of illegal speech in Germany reflects a shift in social discourse on the Internet, and the increasing spread of illegal hate speech, particularly on social media. In 2015, the German Federal Ministry of Justice and Consumer Protection set up a taskforce that included social networks and civil society representatives (Leisegang, 2017[40]). Companies on the taskforce committed to improvements in reporting mechanisms, review, and takedown times for illegal content. Voluntary commitments led to some improvements but the Government considered that more action was required.

The desire for stronger measures led to the making of the Netzwerkdurchsetzungsgesetz (NetzDG) by the Bundestag in June 2017, aimed at addressing hate speech and other illegal content being disseminated online, including terrorist and extreme violent material (Deutscher Bundestag, 2017[41]). The NetzDG considers hate speech and violent content to be illegal only if its circulation and dissemination is subject to criminal prosecution under German Criminal Law. Under NetzDG, Internet platforms with more than 2 million users are required to have reporting systems for hateful posts and to delete reported content that is found to violate one of the 22 relevant statutes of the German Criminal Code. The Act took partial effect in October 2017, and came into full effect on 1 January 2018.

Internet platforms must delete ‘manifestly unlawful’ content within 24 hours of being notified of a complaint. If the content is not obviously illegal, platforms have up to 7 days to make a decision. If a platform receives more than 100 complaints about unlawful content per year, they must publish a transparency report in German every 6 months. Penalties of up to EUR 5 million for individuals, and EUR 50 million for companies for ‘repeated neglect’ apply. Citizens can report violations to Germany’s Federal Office of Justice.

Thus far, only Facebook has been fined under the NetzDG, on grounds of having provided incomplete information in its transparency report for the first half of 2018 on the number of complaints received about unlawful content. According to the German Federal Office of Justice, Facebook’s reporting provided the general public with a distorted image both of the amount of unlawful content and of the social network’s response (German Federal Office of Justice, 2109[42]).

Korea

Korea has passed several anti-terrorism laws that cover online material. Korean legislation allows the head of a related agency to request the cooperation of the head of a ‘relevant institution’ to eliminate, suspend and monitor suspected terrorist or violent extremist content.

In July 2016, the UN General Assembly adopted a resolution calling upon all UN Member States to develop a national plan of action to prevent violent extremism. Accordingly, the government of Korea developed a government-wide plan for preventing violent extremism. The “National Plan of Action for Preventing Violent Extremism” was passed at the National Counter-Terrorism Committee in January 2018 and submitted to

the UN. It includes plans to strengthen public-private cooperation for building a sound Internet environment and to prevent misuse of Internet and communications technologies by terrorist groups.

The Korean government is also participating in the Tech Against Terrorism Initiative led by the UN Counter-Terrorism Executive Directorate (CTED), which uses voluntary contributions for counter-terrorism and operating a Knowledge Sharing Platform for counter-terrorism. The Knowledge Sharing Platform serves as an online knowledge sharing hub that allows large enterprises to transfer their know-how about tackling the misuse of the internet by violent extremist groups to small- and medium-sized IT enterprises.

United Kingdom

The United Kingdom's Terrorist Act 2006 is used to define terrorist content online. This includes considering whether the content seeks to encourage terrorism and the dissemination of terrorist material. The UK government also changed the law through the Counter-Terrorism and Border Security Act 2019, so that people who view terrorist content online could face up to 15 years in prison. This change strengthened the existing offence, so that it applies to material that is viewed or streamed online.

In October 2017 the *Internet Safety Strategy Green Paper* (Bradley MP, 2018[43]) was released, which delivered a vision for a strategic and coordinated approach to online safety and discussed potential actions to address a range of online harms including harassment, trolling, cyberbullying, sexting and online abuse. In May 2018, the UK government published its response to the Green Paper.

The *Online Harms White Paper* (HM Government, 2019[44]) (the White Paper) was published by the UK government's Department for Digital, Culture, Media and Sport (DCMS) and the Home Office on 8 April 2019. The White Paper aims to address a wide scope of online harms, setting out plans for a new system of accountability and oversight for companies.

Central to this approach is the UK government's intention to establish a new statutory duty of care. The duty of care will require companies to take more responsibility for harmful content and behaviour occurring on their platforms. They will need to ensure that they have effective systems and processes in place for reducing and responding to online harm. An independent regulator will be tasked with overseeing compliance with this duty of care.

The White Paper proposes that the regulator should have powers to enable it to:

a. Take enforcement action against companies that do not comply with the duty of care.
b. Establish codes of practice that will set out the steps that companies should take to fulfil their duty of care.
c. Require annual transparency reports from companies, and require additional information from companies to inform its oversight and enforcement activity.
d. Drive improvements to companies' complaints and reporting mechanisms to ensure that they are effective and easy to use.

Other initiatives or proposals canvassed by the White Paper include that interim codes of practice be issued to provide guidance on addressing child sexual exploitation and abuse (CSEA) and terrorist content online in the interim period between the consultation on the White Paper and the regulator being established. An initial consultation response published on 12 February 2020 summarised findings from the White Paper consultation and announced further policy detail, including that the government was minded to appoint Ofcom as the Online Harms regulator. The government will publish the interim codes of practice on addressing child sexual exploitation and abuse (CSEA) and terrorist content online alongside a full consultation response.

United States

The United States approach to TVEC online is guided principally by the First Amendment to the U.S. Constitution which reads, "Congress shall make no law…abridging the freedom of speech." In general, the First Amendment protects a wide range of speech—even speech that is abhorrent or offensive—and generally prohibits prior restraint or censorship of speech by the government. The government may, however, prohibit speech that is directed at inciting or producing imminent lawless action and is likely to incite or produce such action. Therefore, instead of criminalising hateful or abhorrent speech and speech that incites violence or advocates for dangerous causes or groups, the United States has focused on prosecuting criminal activities in furtherance of violence and on promoting credible alternative narratives as the primary means to undermine and counter terrorist messaging.

A number of U.S. statutes criminalise speech-related conduct that supports violent actions, including terrorist acts. For example, under 18 U.S.C. § 373, it is a crime to solicit, command, induce, or otherwise endeavor to persuade another person to engage in a felony involving the threatened, attempted, or actual use of physical force against another person or property, in violation of the laws of the United States.

Additionally, the material support to foreign terrorist organizations statute, 18 U.S.C. § 2339B, applies to actions made under the direction of, or in coordination with, designated foreign terrorist organizations that the actor knows to be terrorist organizations.

Under U.S. law, online service providers are generally protected from liability for the speech of their users, and are protected from liability for their content moderation decisions, except in limited circumstances, including for violations of federal criminal law (see Section 230 of the Communications Decency Act). The U.S. intermediary liability framework facilitates the ability of online service providers to moderate the use of their platforms for types of speech that could not be banned by the government.

Additionally, service providers are prohibited from divulging the contents of electronic communications to the government without user consent, except in certain circumstances (see Stored Communications Act).

Annex A. Global Top 50 Most Popular Online Content-Sharing Services

Rank	Name of service (parent company)	Monthly active users, user accounts or unique visitors (millions)	Type of service	Issues TVEC transparency reports	Provided feedback / comments on its profile
1	Facebook (Facebook, Inc.)	2,320 (as of January 2019) (Kemp, 2019[45])	Social networking and video streaming platform	Y	Y
2	YouTube (Alphabet, Inc.)	1,900 (as of January 2019) (Kemp, 2019[45])	Video streaming platform	Y	Y
3	WhatsApp (Facebook, Inc.)	1,600 (as of January 2019) (Kemp, 2019[45])	Messaging app	N	N
4	Facebook Messenger (Facebook, Inc.)	1,300 (as of January 2019) (Kemp, 2019[45])	Messaging app	N	N
5	iMessage/FaceTime (Apple, Inc)	1,300 (as of January 2019) (Elmer-Dewitt, 2019[46])	Messaging and video chat apps	N	N
6	Weixin/WeChat (Tencent Holdings Ltd.)	1,098 (as of January 2019) (Kemp, 2019[47])	Social networking/content-sharing/messaging platform	N	N
7	Instagram (Facebook, Inc.)	1,000 (as of January 2019) (Kemp, 2019[47])	Social networking platform	Y	Y
8	QQ (Tencent Holdings Ltd.)	807 (as of January 2019) (Kemp, 2019[47])	Instant messaging and web portal site	N	N
9	Youku Tudou (Alibaba Group Holding Limited)	580 (as of August 2019) (Youku Tudou Inc. (NYSE: YOKU), n.d.[48])	Video streaming platform (user-generated and syndicated content)	N	N

10	QZone (Tencent Holdings Ltd.)	531 (as of January 2019) (Kemp, 2019[47])	Social networking platform	N	N
11	Tik Tok (ByteDance Technology Co.)	500 (as of January 2019) (Kemp, 2019[47])	Short video app	N	Y
12	Weibo (Sina Corp.)	462 (as of January 2019) (Kemp, 2019[47])	Social networking platform	N	N
13	iQIYI (Baidu, Inc.)	454 (as of December 2018) (Baidu, Inc., 2018[49])	Video streaming platform (user-generated and syndicated content)	N	N
14	Reddit (Reddit, Inc.)	430 (as of October 2019) (Murphy, 2019[50])	Social news aggregation, web content ranking and discussion website	N	Y
15	Twitter (Twitter, Inc.)	326 (as of January 2019) (Kemp, 2019[47])	Short messages-focused social networking platform	Y	Y
16	Douban (Information Technology Company, Inc.)	320 (as of January 2019) (Kemp, 2019[47])	Social networking platform	N	N
17	LinkedIn (Microsoft, Inc.)	303 (as of January 2019) (Kemp, 2019[47])	Jobs-focused social networking platform	N*	N
18	Baidu Tieba (Baidu, Inc.)	300 (as of January 2019) (Kemp, 2019[47])	Online communications platform	N	N
19	Skype (Microsoft, Inc.)	300 (as of January 2019) (Kemp, 2019[47])	Video chat and voice calls app	N*	N
20	Quora (Quora, Inc.)	300 (as of September 2018) (Marketing Land, 2018[51])	Question-and-answer website	N	N
21	Snapchat (Snap, Inc.)	287 (as of January 2019) (Kemp, 2019[47])	Social networking platform	N	Y
22	Viber (Rakuten, Inc.)	260 (as of January 2019) (Kemp, 2019[47])	Messaging app	N	Y
23	Pinterest (Pinterest, Inc.)	250 (as of January 2019) (Kemp,	Social networking platform	N	Y

		2019[47])			
24	Vimeo (Vimeo, Inc.)	240 (as of September 2018) (Bicknell, 2018[52])	Video streaming app	N	Y
25	IMO (PageBites, Inc.)	211 (as of April 2019) (YY Inc. - IR Site, 2019[53])	Video chat and voice calls app	N	N
26	Telegram (Telegram Messenger LLP)	200 (as of March 2018) (Pavel, 2018[54])	Messaging app	N	N
27	LINE (Line Corporation)	194 (as of January 2019) (Kemp, 2019[47])	Messaging app	N	Y
28	Ask.fm (IAC [InterActiveCorp])	160 (as of August 2018) (Kallas, 2019[55])	Social networking platform	N	Y
29	Twitch (Amazon.com, Inc.)	140 (as of February 2019) (Iqbal, 2019[56])	Livestreaming platform	N*	Y
30	Xigua (ByteDance Technology Co.)	121 (as of December 2018) (Yang, 2019[57])	Short video streaming app	N	N
31	Tumblr (Automattic, Inc.).	115 (as of August 2018) (Kallas, 2019[55])	Microblogging and social networking platform	N	N
32	Flickr (SmugMug, Inc.)	112 (as of August 2018) (Kallas, 2019[55])	Image and video hosting service	N	N
33	Huoshan (ByteDance Technology Co.)	99 (as of December 2018) (Yang, 2019[57])	Short video streaming app	N	N
34	VK (Mail.Ru Group)	97 (as of August 2018) (Kallas, 2019[55])	Social networking platform	N	Y
35	YY Live/Huya (YY, Inc.)	90 (as of December 2018) (Baidu, Inc., 2018[49])	Livestreaming platform	N	N
36	Medium (A Medium Corporation.)	86 (as of August 2018) (Wickey, 2018[58])	Online publishing platform	N	Y
37	Haokan (Baidu, Inc.)	75 (as of December	Short video streaming	N	N

		2018) (Yang, 2019[57])	app		
38	Odnoklassniki (Mail.Ru Group)	71 (as of August 2018) (Kallas, 2019[55])	Social networking platform	N	N
39	Discord (Discord, Inc.)	56 (as of May 2019) (Vincent, 2019[59])	Chat platform	N	N
40	Smule (Smule, Inc.)	52 (as of July 2018) (Solsman, 2018[60])	User-generated music-video sharing platform	N	N
41	KaoKao Talk (Daum Kakao Corporation)	50 (as of January 2019) (Statista, 2019[61])	Messaging app	N	Y
42	Deviantart (DeviantArt, Inc.)	45 (as of 2016) (DeviantArt Media Kit, n.d.[62])	Online artwork, videography and photography platform	N	N
43	Meetup (WeWork Companies, Inc.)	35 (as of August 2018) (Kallas, 2019[55])	Interest-based social networking platform	N	N
44	4chan (4chan Community Support LLC)	22 (as of August 2019) (4chan, n.d.[63])	Content-sharing platform	N	N
45	MySpace (Viant Technology/Meredith Corporation)	15 (as of April 2016) (Barr, 2016[64])	Music-oriented social networking platform	N	N

* On 15 May 2019, in connection with the Christchurch Call, Amazon, Facebook, Google, Microsoft and Twitter committed to "publishing on a regular basis transparency reports regarding detection and removal of terrorist or violent extremist content" on their platforms and services, and to "ensuring that the data is supported by a reasonable and explainable methodology" (https://blogs.microsoft.com/on-the-issues/2019/05/15/the-christchurch-call-and-steps-to-tackle-terrorist-and-violent-extremist-content/). Amazon and Microsoft have not issued TVEC-specific transparency reports yet, though.

Monthly active user (MAU) data are unavailable for certain other online content-sharing services that terrorists and violent extremists have used, yet the metrics that are available suggest that they should be included in the top 50 list. The table therefore continues below with five more services, but without ranks because metrics other than MAU indicate their significance, so a proper comparison with the services above was not possible. In any event, for purposes of this report, the overall composition of the group of 50 is more important than the individual rankings.

Name of service (parent company)	Indicative Global Market Share	Type of market/service	Transparency report on terrorist/violent extremist content	Provided feedback / comments on its profile
Google Drive (Alphabet, Inc.)	34.63% (as of October 2019) (Datanyze, 2019[65])	Cloud-based file sharing	N	N
Dropbox (Dropbox, Inc.)	24.08% (as of October 2019) (Datanyze, 2019[65])	Cloud-based file sharing	N	N
Microsoft OneDrive (Microsoft, Inc.)	10.95% (as of October 2019) (Datanyze, 2019[65])	Cloud-based file sharing	N	N

Name of service (parent company)	Indicative Global Market Share or monthly pageviews	Type of market/service	Transparency report on terrorist/violent extremist content	Provided feedback / comments on its profile
Wordpress.com (Automattic, Inc.)	60% (as of April 2019) (Kinsta, 2011-2019[66])	Content management system	Y	N
Wikipedia (Wikimedia Foundation)	18 billion pageviews per month (as of January 2016) (Pew Research Center, 2016[67]); 10th most visited website worldwide (Alexa, 2019[68])	Online encyclopaedia	N	N

Annex B. Profiles of the Top 50 Services

1. Facebook[1]

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition of TVEC. However, Facebook is one of the few Services with a well-developed definition of terrorism and related terms. In the section of Facebook's Community Standards entitled 'Dangerous Individuals and Organisations' (Facebook, n.d.[69]), Facebook states that any organisations or individuals that proclaim a violent mission or are engaged in violence cannot have a presence on Facebook. Such organisations or individuals are defined to include those involved in: ● Terrorist activity ● Organised hate ● Mass murder (including attempts) or multiple murder ● Human trafficking ● Organized violence or criminal activity Content that expresses support or praise for groups, leaders or individuals involved in these activities is removed. Also, the following people (whether living or deceased) and groups cannot maintain a presence (for example, have an account, Page or group) on Facebook: terrorist organisations, terrorists, hate organisations (and their leaders and prominent members) and mass and multiple murderers. Terrorist organisations and terrorists include any non-state actor that: ● Engages in, advocates or lends substantial support to purposive and planned acts of violence, ● Which causes or attempts to cause death, injury or serious harm to civilians, or any other person not taking direct part in the hostilities in a situation of armed conflict, and/or significant damage to property linked to death, serious injury or serious harm to civilians ● With the intent to coerce, intimidate and/or influence a civilian population, government or international organisation ● In order to achieve a political, religious or ideological aim. A hate organisation is defined as any association of three or more people that is organised under a name, sign or symbol and that has an ideology, statements or physical actions that

	attack individuals based on characteristics, including race, religious affiliation, nationality, ethnicity, gender, sex, sexual orientation, serious disease or disability. A homicide is considered to be a mass murder if it results in three or more deaths in one incident. Any individual who has committed two or more murders over multiple incidents or locations is deemed a multiple murderer. Facebook prohibits any symbols that represent any of the above organisations or individuals, unless they are shared with context that condemns or neutrally discusses the content. Content that praises any of the above organisations or individuals or any acts committed by them is prohibited. Also, Facebook does not allow coordination of support for any of the above organisations or individuals or any acts committed by them. Further, Facebook prohibits content that represents or supports in any way events that it designates as terrorist attacks, hate crimes, or mass shootings. Lastly, in the section titled 'Violence and Incitement' of Facebook's Community Standards (Facebook, n.d.[70]), Facebook states that it removes language that incites or facilitates serious violence. In particular, users cannot post: ● Threats that could lead to death (and other forms of high-severity violence) of any target(s), where threat is defined as any of the following: ○ Statements of intent to commit high-severity violence ○ Calls for high-severity violence including content where no target is specified but a symbol represents the target and/or includes a visual of an armament to represent violence; or ○ Statements advocating for high-severity violence; or ○ Aspirational or conditional statements to commit high-severity violence ● Content that asks or offers services for hire to kill others (for example, hitmen, mercenaries, assassins) or advocates for the use of a hitman, mercenary or assassin against a target. ● Admissions, statements of intent or advocacy, calls to action or aspirational or conditional statements to kidnap a target. ● Threats that lead to serious injury (mid-severity violence) towards private individuals, minor public figures, vulnerable persons or vulnerable groups, where threat is defined as any of the following: ○ Statements of intent to commit violence ○ Statements advocating violence; or

	○ Calls for mid-severity violence including content where no target is specified but a symbol represents the target; or ○ Aspirational or conditional statements to commit violence; or ○ Content about other target(s) apart from private individuals, minor public figures, vulnerable persons or vulnerable groups and any credible: ▪ Statements of intent to commit violence; ▪ Calls for action of violence; ▪ Statements advocating for violence; or ▪ Aspirational or conditional statements to commit violence ● Threats that lead to physical harm (or other forms of lower-severity violence) towards private individuals (self-reporting required) or minor public figures, where threat is defined as any of the following: ○ Statements of intent ○ calls for action ○ advocating, aspirational, or conditional statements to commit low-severity violence ● Imagery of private individuals or minor public figures that has been manipulated to include threats of violence either in text or pictorially (adding bullseye, dart, gun to head etc.) ● Any content created for the express purpose of outing an individual as a member of a designated and recognisable at-risk group ● Instructions on how to make or use weapons if there is evidence of a goal to seriously injure or kill people, through: ○ Language explicitly stating that goal, or ○ photos or videos that show or simulate the end result (serious injury or death) as part of the instruction, ○ unless the aforementioned content is shared as part of recreational self-defence, for military training purposes, commercial video games or news coverage (posted by Page or with news logo) ● Providing instructions on how to make or use explosives, unless there is clear context that the content is for a non-violent purpose (for example, part of commercial video games, clear scientific/educational purpose, fireworks or specifically for fishing) ● Any content containing statements of intent, calls for action or advocating for high or mid-severity violence

	due to voting, voter registration or the outcome of an election ● Misinformation that contributes to imminent violence or physical harm; and ● Calls to action, statements of intent to bring armaments to locations, including but not limited to places of worship, or encouraging others to do the same.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.facebook.com/communitystandards/.
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	Yes, available at https://about.fb.com/news/2019/05/protecting-live-from-abuse/. In particular, Facebook applies a 'one strike' policy to prohibited livestreamed content, meaning that anyone who violates Facebook's 'most serious policies' will be restricted from using Live for set periods of time, for example 30 days, starting on their first offense.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Facebook removes content from the platform when content violates its Community Standards.
4.1 Notifications of removals or other enforcement decisions	After the content removal, the person who posted the content is notified and given the option to request a review or accept the decision (Facebook, n.d.[71]).
4.2 Appeal processes against removals or other enforcement decisions	If the user requests a review, the content is resubmitted for another review. The content is not visible to other people on Facebook while under review. Reviewers do not know that the post has been reviewed previously. It is not clear, based on the Community Standards, whether the review is done by a single person or a panel of people, or what training or qualifications the reviewers have. If the reviewer agrees with the original decision, the content remains off Facebook. However, if the reviewer disagrees with the initial review and decides it should not have been removed, the content will go to a third reviewer. This reviewer's decision will determine whether the content is allowed on Facebook or not. For some violation types (which are not specified), Facebook also allows the person who posted to request a review a second time. In this second round, the content is reviewed by reviewers who are experts on that particular violation type, and the person appealing has the opportunity to provide more information in a text field (Facebook, n.d.[71]).

5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Facebook is careful to balance transparency with pragmatism regarding the amount of information they share with the public, to avoid giving terrorists the tools to circumvent their enforcement techniques (Facebook, 2018[72]). However, it has stated that it removes 'terrorists and posts that support terrorism' whenever it becomes aware of them. (Facebook, 2017[73]) Also, when Facebook receives reports of potential terrorism posts, it reviews those reports urgently and with scrutiny, and in the rare cases when it uncovers evidence of imminent harm, Facebook promptly informs the authorities (Facebook, 2017[73]). Facebook uses artificial intelligence (AI) as one of its tools to combat terrorism, including techniques such as image matching, language understanding, removal of terrorist clusters and cross-platform collaboration (i.e. with WhatsApp and Instagram). Recently, Facebook started using machine learning to assess Facebook posts that may signal support for ISIS or al-Qaeda (Facebook, 2018[72]). Furthermore, Facebook notes that AI cannot catch everything, so it also relies on human expertise, including Facebook users (who may report terrorist-related content), its 'Community Operations team' (Facebook, n.d.[71]), terrorism and safety specialists, cooperation with other tech firms such as Microsoft, Twitter and YouTube, and government and inter-governmental agencies. Facebook also reports that it supports counterspeech programs such as the Online Civil Courage Initiative (Facebook, 2017[73]). The marginal economic costs of using AI tools to identify TVEC are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. Facebook is a founding member of GIFCT and participates in its Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	The consequences for breaching Facebook's Community Standards vary depending on the severity of the breach and a person's history on the platform. Prohibited content may be removed. In addition, Facebook may warn someone after a first breach, but if the user continues to breach Facebook's policies, Facebook may restrict the user's ability to post on Facebook or disable their profile. Facebook may also notify law enforcement when it believes that there is a genuine risk of physical harm or a direct threat to public safety.
7. Does the service issue transparency reports (TRs) specifically on content related to terrorism and/or violent extremism?	Yes (Facebook, 2018-2019[74]). Facebook issues transparency reports on the enforcement of its Community Standards, and one section is about 'Terrorist Propaganda' while another is about 'Violence and Graphic Content'. Note that Facebook states that it does not tolerate any content that praises, endorses or represents terrorist organisations or terrorists. Facebook enforces this standard as applied to terrorist activities and groups both regionally and globally. Since

	November 2019, its terrorist propaganda TRs measure the actions Facebook takes against all terrorist organisations, rather than focusing just on propaganda related to ISIS, al-Qaeda and their affiliate groups (Facebook, 2020[7]).
8. What information/fields of data are included in the TRs?	The latest report, issued in November 2019, includes the following five fields of information in both the 'Terrorist Propaganda' section and the 'Violence and Graphic Content' section: - *Prevalence (How prevalent were terrorist propaganda, and violence and graphic content, violations on Facebook?)* The prevalence metric is the percentage of all views that were of content that violated certain of Facebook's community standards. For example, Facebook estimates that less than 0.04% of views were of content that violated its standards for terrorist propaganda in Q3 2019. In other words, fewer than 4 of every 10,000 views on Facebook contained what the company deemed to be terrorist propaganda. (The figures refer to final determinations, not content that was initially flagged as a possible violation but may have been subsequently determined to be permissible.) - *Content actioned (How much content did Facebook take action on?)* Facebook indicates that a piece of content can be 'any number of things', (Facebook, n.d.[71]) including a post, photo, video or comment. Taking action may include removing a piece of content from Facebook, covering photos or videos that may be disturbing to some audiences with a warning, or disabling accounts. Content actioned is the total number of pieces of content that Facebook took action on during a given reporting period because it violated its community standards. - *Proactive rate (Of the violating content actioned, how much did Facebook find before users reported it?)* This metric shows the percentage of content actioned for violating Facebook's policies that Facebook found and flagged before users reported it. It counts detections made by both Facebook's AI tools and human reviewers. - *Appeals (How much of the content Facebook actioned did people appeal?)* This metric counts the number of pieces of content actioned for which people requested another review during the reporting period. - *Restored content (How much content did Facebook restore after removing it?)* Restored content is the number of pieces of content that Facebook restored during the reporting period after previously actioning it.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	- *Prevalence.* The prevalence metric is the estimated number of views of violating content, divided by the estimated number of total content views on Facebook, per reporting period. For example, if the prevalence of terrorist propaganda is 0.18% to 0.20%, that means of every 10,000 content views, 18 to 20 on average were

	of content that violated Facebook's standards for terrorist propaganda. The prevalence metric provides an indication of how often prohibited content is seen, rather than the total amount of such content published. Prevalence is estimated based on samples of content across different areas of Facebook, such as Groups and News Feeds. For terrorist propaganda violations, in particular, Facebook only estimates the upper limit, which means that Facebook is 'confident that the prevalence of violating views is below that limit.' (Facebook, n.d.[71]) Content on both Facebook and Messenger are included in this metric. - *Content actioned.* Content actioned is the total number of pieces of content that Facebook took action on during a given reporting period because it violated its content policies. Facebook does not count those scenarios where it escalates content to law enforcement. This metric includes both content Facebook actioned after someone reported it and content that Facebook found proactively. - *Proactive rate.* This metric is calculated as: the number of pieces of content actioned that Facebook found and flagged before users reported them, divided by the total number of pieces of content actioned. Content on Facebook and Messenger are included in this metric. - *Appeals.* This metric counts the number of pieces of content actioned for which people requested another review during the reporting period. Content on Facebook and Messenger are included in this metric. - *Restored content.* To arrive at this metric, Facebook counts the number of pieces of content that it restored during the reporting period after previously actioning it. Facebook may restore content either when a decision to remove is appealed or when Facebook discovers a reason to restore the content. Only Facebook content is included in this metric.
10. Frequency/timing with which TRs are issued	Facebook indicates that it issues transparency reports 'regularly'. Facebook has issued 4 3 TRs. One for the period Q4 2017 – Q1 2018; one for Q2 2018 – Q3 2018; another for Q4 2018 – Q1 2019; and the last for Q2 2019 – Q3 2019.
11. Has this service been used to post TVEC?	Yes. See above sections 7-9.

2. YouTube

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition of TVEC. However, YouTube's Community Guidelines contain a number of clarifications that are relevant to terrorist and violent extremist content. The policy on Violent Criminal Organisations, for example, states that content intended to praise, promote, or aid violent criminal organisations is not allowed on YouTube. In addition, such organizations are banned from YouTube for any purpose, including recruitment. The Guidelines neither contain nor

	refer to a list of such organisations, though. Nevertheless, the policy prohibits the following types of content: • Content produced by violent criminal or terrorist organisations • Content praising or memorialising prominent terrorist or criminal figures in order to encourage others to carry out acts of violence • Content praising or justifying violent acts carried out by violent criminal or terrorist organisations • Content aimed at recruiting new members to violent criminal or terrorist organisations • Content depicting hostages or posted with the intent to solicit, threaten, or intimidate on behalf of a violent criminal or terrorist organisation • Content that depicts the insignia, logos, or symbols of violent criminal or terrorist organisations in order to praise or promote them. If content related to terrorism or crime is posted for an educational, documentary, scientific, or artistic purpose, enough information in the video or audio must be included so viewers understand the context. The policy on Violent Criminal Organisations also gives the following examples of content that is not allowed on YouTube: • Raw and unmodified reuploads of content created by terrorist or criminal organisations • Celebrating terrorist leaders or their crimes in songs or memorials • Celebrating terrorist or criminal organisations in songs or memorials • Content directing users to sites that espouse terrorist ideology, are used to disseminate prohibited content, or are used for recruitment. Moreover, YouTube's violent or graphic content policies prohibits violent or gory content intended to shock or disgust viewers, or content encouraging others to commit violent acts. In particular, it prohibits the following types of content: • Inciting others to commit violent acts against individuals or a defined group of people. • Footage, audio or imagery involving road accidents, natural disasters, war aftermath, terrorist attack aftermath, street fights, physical attacks, sexual assaults, immolation, torture, corpses, protests or riots, robberies, medical procedures or other such scenarios with the intent to shock or disgust viewers. YouTube's policy on hate speech bans content promoting violence or hatred against individuals or groups based on any of the following attributes: Age, Caste, Disability, Ethnicity, Gender Identity, Nationality, Race, Immigration Status, Religion, Sex/Gender, Sexual Orientation, Victims of a major violent event and their kin, and Veteran Status.

	Content that encourage violence against individuals or groups based on any of on the attributes noted above, or that incites hatred against individuals or groups based on any of the attributes noted above, is prohibited. Among the examples provided of content that falls within this category is praising or glorifying violence against individuals or groups based on the attributes noted above. Lastly, the policy on harmful or dangerous content bans instructions to kill or harm. This means showing viewers how to perform activities meant to kill or maim others, such as providing instructions on how to build a bomb meant to injure or kill people. Also prohibited is content about violent events if it promotes or glorifies violent tragedies such as school shootings.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	YouTube's Community Guidelines are available at https://www.youtube.com/about/policies/#community-guidelines Guidelines on Violent Criminal Organisations are available at https://support.google.com/youtube/answer/9229472?hl=en&ref_topic=9282436 Guidelines on violent or graphic content are available at https://support.google.com/youtube/answer/2802008?hl=en-GB&ref_topic=9282436 Guidelines on hate speech are available at https://support.google.com/youtube/answer/2801939?hl=en Guidelines on harmful or dangerous content are available at https://support.google.com/youtube/answer/2801964?hl=en&ref_topic=9282436
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No. YouTube's Community Guidelines apply to videos, video descriptions, comments, live streams and any other YouTube product or feature.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	If content violates any of YouTube's content policies, YouTube removes the content.
4.1 Notifications of removals or other enforcement decisions	The content removal is notified to users via email, desktop or mobile notifications, and an alert in their channel settings (Google/Youtube, 2020[75]). If the content removal results in a 'strike' (see below section 6), YouTube informs the user: • What content was removed • Which policies it violated • How the strike affects the user's channel • What the user can do next

4.2 Appeal processes against removals or other enforcement decisions	When users receive a strike, and they believe YouTube made a mistake, they can appeal the strike (Google, Youtube, 2020[76]). YouTube informs users about the result of the appeal via email. The result may be any of the following: • If YouTube finds that the content followed YouTube's Community Guidelines, YouTube reinstates it and removes the strike from the user's channel. If the user appeals a warning (see below section 6) and the appeal is granted, the next offense will result in a warning. • If YouTube finds that the content followed YouTube's Community Guidelines, but is not appropriate for all audiences, an age-restriction is applied. If the content is a video, it will not be visible to users who are signed out, are under 18 years of age, or have Restricted Mode (Google, Youtube, 2020[77]) turned on. If the content is a custom thumbnail, it will be removed. • If YouTube finds that the content was in violation of YouTube's Community Guidelines, the strike will stay and the video will remain off the platform. There is no additional penalty for appeals that are rejected. Users may appeal each strike only once.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	YouTube provides its users with tools to report content that violates its Community Guidelines (Google, Youtube, 2020[78]). YouTube has also developed automated systems that aid in the detection of content that may violate its policies. When its automated systems flag potentially problematic content, human reviewers then verify whether it indeed violates company policies. If it does, the content is removed and is used to train YouTube's automated systems to perform better in the future. With respect to the automated systems that detect extremist content (an undefined term) in particular, YouTube's staff have manually reviewed over two million videos to provide training examples. In addition, YouTube invests in a network of over 180 academics, government partners and NGOs who bring expertise to the platform's enforcement systems, including through YouTube's Trusted Flagger programme. (Google, Youtube, 2020[79])[2] In the context of violent extremism, this includes the International Centre for the Study of Radicalisation at King's College, London (The International Centre for the Study of Radicalisation (ICSR), 2020[80]), the Institute for Strategic Dialogue (ISDGlobal, n.d.[81]), the Wahid Institute in Indonesia and government agencies focused on counterterrorism. Participants in the Trusted Flagger programme receive training in enforcing YouTube's Community Guidelines, and because their flags have a higher action rate than the average user, YouTube prioritises them for review. Otherwise, content flagged by Trusted Flaggers is subject to the same policies as content flagged by any other user and is reviewed by teams that are trained to make decisions on whether content violates YouTube's Community Guidelines. Individual users, government agencies, and NGOs are eligible for participation in the YouTube Trusted Flagger programme. Participants must be committed to frequently flagging content that may violate

	YouTube's Community Guidelines and be open to ongoing discussion and feedback on various YouTube content areas. YouTube notes that hate speech is a complex policy area to enforce at scale, as decisions require nuanced understanding of local languages and contexts. For consistent enforcement of its hate speech policy, YouTube has expanded its review team's linguistic and subject matter expertise. YouTube also deploys machine learning to better detect potentially hateful content to send for human review, applying lessons from its enforcement against other types of content, like violent extremism (Google, Youtube, n.d.[82]). The marginal economic costs of using automated tools to identify TVEC are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. YouTube is a founding member of GIFCT and participates in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	The first time a user posts content that violates YouTube's Community Guidelines, he or she receives a warning with no penalty to their channel. For subsequent violations, YouTube issues a 'strike' against the user's channel. The channel is terminated if the user receives 3 strikes within a 90-day period. When the first strike is issued, the user cannot do any of the following for one week: • Upload videos, live streams, or stories • Create custom thumbnails or Community posts • Created, edit, or add collaborators to playlists • Add or remove playlists from the watch page using the "Save" button Full privileges are restored automatically after the 1-week period, but the strike will remain on the user's channel for 90 days. If the user gets a second strike within 90-days of the first strike, the user will not be able to post content for two weeks. If there are no further issues, full privileges are restored automatically after the 2-week period, but each strike expires 90 days from the time it was issued. Three strikes in the same 90-day period will result in the user's channel being permanently removed from YouTube (Google, YouTube, n.d.[83]). Beyond the three strikes system, a YouTube channel will be terminated if it has a single case of severe abuse (such as predatory behaviour) or is determined to be wholly dedicated to violating YouTube's guidelines (as is often the case with spam accounts). When a channel is terminated, all of its videos are removed. Content that does not violate YouTube's policies but is close to meeting the criteria for removal and could be offensive to some viewers may have some features disabled. This may include the following:

	• Inflammatory religious or supremacist content without a direct call to violence or a primary purpose of promoting hatred • Conspiracy theories ascribing evil, corrupt or malicious intent to individuals or groups based on certain attributes • Videos denying that a well-documented violent event took place The content will remain available on YouTube, but the watch page will no longer have comments, suggested videos or likes, and will be placed behind a warning message. These videos are also not eligible for ads. Having features disabled will not add a strike to the video owner's channel (Google, YouTube, n.d.[84]). YouTube notifies decisions to disable features via email. Users can appeal this decision.
7. Does the service issue transparency reports (TRs) on TVEC?	Yes (Google, n.d.[85]). YouTube issues transparency reports on the enforcement of its Community Guidelines. One section of these reports is about 'Violent Extremism' (Google, YouTube, n.d.[86]). The last TR specifies that content that violates YouTube's policies against violent extremism includes material produced by government-listed foreign terrorist organisations (YouTube does not specify which government(s) it is referring to, though). The TR also specifies that YouTube strictly prohibits content that promotes terrorism, such as content that glorifies terrorist acts or incites violence. In addition, the TR states that content produced by violent extremist groups that are not government-listed foreign terrorist organisations is often covered by YouTube's policies against posting hateful or violent or graphic content (see Section 1 above), including content that is primarily intended to be shocking, sensational or gratuitous.
8. What information/fields of data are included in the TRs?	YouTube discloses • the number of content removal requests by governments based on six categories (national security, defamation, regulated goods and services, privacy and security, copyrights and 'all others') (Google, 2010-2019[87]); • the number of channels removed, separated by ground of removal (amongst which are the promotion of violence and violent extremism); • the number of videos removed by source of first detection (automated flagging, individual trusted flagger, users, NGOs and governments); • the percentage of videos first flagged through automated flagging systems, with and without views, i.e. the percentage of removals that occurred before the videos received any views versus those that occurred after the videos received some views; • the number and percentage of human flags, by flagging reason (including the promotion of terrorism). YouTube notes that a video may be flagged multiple times for multiple reasons, and that flagging it does not necessarily result in removal. Human-flagged videos are removed for violations of Community Guidelines once a trained reviewer confirms a policy violation (Google, 2010-2019[87]).

	• the percentage and number of videos removed, by removal reason (including under YouTube's violent extremism policy and hate speech policy) (Google, YouTube, n.d.[86]); • the number of comments removed, by removal reason (including under YouTube's violent extremism policy and hate speech policy); and • the percentage of removed comments by source of first detection (automated flagging and human flagging) (Google, n.d.[85]).
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	No information is provided.
10. Frequency/timing with which TRs are issued	On a quarterly basis (Google, n.d.[88]).
11. Has this service been used to post TVEC?	Yes. See above sections 7-8.

3. WhatsApp

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	WhatsApp's ToS do not define TVEC. However, in the section titled 'Safety and Security' in WhatsApp's ToS states that WhatsApp works to protect the safety and security of WhatsApp by appropriately 'dealing with abusive people and activity' and violations of its Terms. It is possible that the concept 'abusive people and activity' encompasses users disseminating TVEC, although this is not stated explicitly. 'Abusive people and activity' is not defined. The ToS also state that WhatsApp prohibits misuse of its services, 'harmful conduct towards others', and violations of its Terms and policies. WhatsApp notes that users must access and use its services only for 'legal, authorised, and acceptable purposes', which includes not using its services in ways that "are illegal, obscene, defamatory, threatening, intimidating, harassing, hateful, racially or ethnically offensive, or instigate or encourage conduct that would be illegal or otherwise inappropriate, including promoting violent crimes."
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.whatsapp.com/legal/#terms-of-service
3. Are there specific provisions applicable to livestreamed content in	No.

the ToS or Community Guidelines/Standards?	
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	WhatsApp broadly states that it may modify, suspend, or terminate a user's access to or use of its services at any time for suspicious or unlawful conduct, or if it reasonably believes that the user is violating its Terms or creating harm or risk for users or other people. No appeal processes are specified. However, if a user believes that his or her account was terminated or suspended by mistake, the user can contact WhatsApp at support@whatsapp.com.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified. However, if a user believes that his or her account was terminated or suspended by mistake, the user can contact WhatsApp at support@whatsapp.com.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	WhatsApp states that it develops automated systems to improve its ability to detect and remove 'abusive people and activity' that may harm WhatsApp's community and the safety and security of its services. Also, users can report any content they may deem problematic, and WhatsApp's moderators review those reports to take appropriate action. Since WhatsApp is part of the 'Facebook Companies', it is possible that it uses the same methods as Facebook to identify and remove terrorist and violent content, not least because WhatsApp notes that it shares information with the Facebook Companies to fight spam, threats, abuse, or infringement activities and promote safety and security across the Facebook Company Products. However, since WhatsApp communications are encrypted, it is difficult to imagine how any TVEC content can be intercepted. The marginal economic costs of using automated tools to identify TVEC are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. WhatsApp recently became a member of the GIFCT.
6. Sanctions/consequences in case of breaches of ToS or Community Guidelines/Standards	If a user violates WhatsApp's ToS or policies, WhatsApp may take action with respect to the user's account, including disabling or suspending it. If WhatsApp does so, the user must not create another account without WhatsApp's permission. WhatsApp also notes that if it becomes aware of 'abusive people or activity', it will take appropriate action by removing such people or activity or contacting law enforcement.

7. Does the service issue transparency reports (TRs) on TVEC	Not yet, but issuing TRs is a condition of membership in GIFCT, so WhatsApp may be expected to do so in the near future.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. For example, after the Christchurch shootings, two far-right violent extremists reportedly were part of a WhatsApp group called 'Christian White Militia' and published statements encouraging terrorism in March 2019 (Dearden, 2019[89]).

4. Facebook Messenger

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition of TVEC. Facebook Messenger does not have specific ToS or Community Standards. However, as Facebook scans Facebook Messenger conversations to detect violations to its Community Standards, (Frier, 2018[90]) these Standards, which feature a well-developed description of terrorism and related concepts, apply to Facebook Messenger. See Section 1 of the Facebook Profile.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.facebook.com/communitystandards/
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	See Section 4 of the Facebook Profile.

4.1 Notifications of removals or other enforcement decisions	See Section 4.1 of the Facebook Profile.
4.2 Appeal processes against removals or other enforcement decisions	See Section 4.2 of the Facebook Profile.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	See Section 5 of the Facebook Profile.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	See Section 6 of the Facebook Profile.
7. Does the service issue transparency reports (TRs) on TVEC	See Section 7 of the Facebook Profile.
8. What information/fields of data are included in the TRs?	See Section 8 of the Facebook Profile.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	See Section 9 of the Facebook Profile.
10. Frequency/timing with which TRs are issued	See Section 10 of the Facebook Profile.
11. Has this service been used to post TVEC?	Yes. See above sections 7-8 of the Facebook Profile.

5. iMessage/FaceTime

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, Apple's Media Services Terms and Conditions (which govern iMessage and FaceTime) prohibit users from posting objectionable, offensive, unlawful, deceptive or harmful content, such as comments, pictures, videos, and podcasts (including associated metadata and artwork).
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.apple.com/ca/legal/internet-services/itunes/ca/terms.html

3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	No procedures are specified. Apple broadly states that it may monitor and decide to remove or edit any submitted material.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Apple has a reporting mechanism that allow users to report content that violates its Submission Guidelines (included in Apple's Media Services Terms and Conditions). These reports are verified and processed by Apple's team. Given that iMessage and FaceTime are encrypted, it is difficult to see how an algorithm or an on-staff reviewer who works for Apple could detect any problematic content, including TVEC. The marginal economic costs of using human moderators to identify problematic content are probably relatively high. Apple is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	If Apple determines there is a breach or suspected breach of any of the provisions of its ToS, Apple may, without notice to the user, terminate the user's Apple ID, license to Apple's software and/or access to its services, which include iMessage and FaceTime.
7. Does the service issue transparency reports (TRs) on TVEC?	No. Apple does issue transparency reports (Apple, n.d.[91]) that contain a section on content removal requests from governments and private parties reporting violations of its ToS or local laws, but there is no specific information on TVEC.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.

10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Possibly. A security manual issued by ISIS recommended use of iMessage to protect supporters' identities, (Zetter, 2015[92]) but there is no evidence that ISIS supporters have actually used it (Dilger, 2015[93]).

6. WeChat

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no definition. However, in its ToS, WeChat prohibits its users from submitting, uploading, transmitting or displaying any content which in fact or in WeChat's reasonable opinion: • breaches any laws or regulations (or may result in a breach of any laws or regulations); • creates a risk of loss or damage to any person; • harms or exploits any person (whether adult or minor) in any way, including via bullying, harassment or threats of violence; and • is hateful, harassing, abusive, racially or ethnically offensive, defamatory, humiliating to other people (publicly or otherwise), threatening, profane or otherwise objectionable.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.wechat.com/en/service_terms.html and https://www.wechat.com/en/acceptable_use_policy.html (Tencent, n.d.[94])
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	WeChat broadly states that it may review (but make no commitment to review) content (including any content posted by WeChat users) or third party programs or services made available through WeChat to determine whether or not they comply with WeChat's policies, applicable laws and regulations or are otherwise objectionable, and WeChat reserves the right to block or remove content for any reason, as required by applicable laws and regulations.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.

4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	WeChat provides no information in this regard. It has been reported that Chinese online firms, including WeChat, have a team of moderators policing problematic content. [3] Also, research has shown that WeChat uses algorithmic technology (Knockel, 2018[16]), keyword filtering and URL blocking (Ruan, 2016[21]) to censor content that is in violation of its ToS (which may include the posting of TVEC), although these methods are reportedly applied only to accounts registered to mainland China phone numbers (Ruan, 2016[21]). The marginal economic costs of using automated tools to identify problematic content are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. WeChat is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	WeChat notes that it may suspend or terminate access to WeChat if it reasonably believes that a user has breached WeChat's ToS, their use of WeChat creates risk for WeChat or other WeChat users, the suspension or termination is required by applicable laws, or at WeChat's sole and absolute discretion.
7. Does the service issue transparency reports (TRs) on TVEC	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. The Christchurch shooting was posted on WeChat (Kenny, 2019[95]). In addition, WeChat has been used to disseminate anti-Muslim propaganda (Huang, 2018[96]).

7. Instagram

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Facebook and Instagram share policies, generally. Facebook notes that if content is considered to be in violation of such policies on Facebook, it would also be in violation on Instagram. Therefore, Instagram follows the definitions set forth in Facebook's profile (see Section 1 of Facebook's profile). Because Facebook's Community Standards are more comprehensive than Instagram's Community Guidelines, they are the point of reference, even when considering Instagram violations Instagram's Community Guidelines provide that Instagram is not a place to support or praise terrorism, organized crime, or hate groups, or to encourage violence or attack anyone based on their race, ethnicity, national origin, sex, gender, gender identity, sexual orientation, religious affiliation, disabilities, or diseases. Also, serious threats of harm to public and personal safety are prohibited, as well as the sharing of graphic images to glorify violence.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Instagram's Community Guidelines are available at https://help.instagram.com/477434105621119?helpref=page_content Instagram's ToS are available at https://help.instagram.com/581066165581870
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Instagram may remove content if it violates its Community Guidelines, or it may disable or terminate an account.
4.1 Notifications of removals or other enforcement decisions	Instagram notifies the affected user of such content removals or account suspension or termination.
4.2 Appeal processes against removals or other enforcement decisions	If users believe their content has been removed or their account has been terminated in error, they can appeal the decision. It is possible for users to appeal the removal of content that was deemed to violate Instagram's 'counter-terrorism' policies (which are not specified). If content is found to have been removed in error, Instagram will restore the post and remove the violation from the account's record. In February 2020, Instagram rolled out a streamlined appeals process for disabled accounts directly through the app, instead of

	through the Instagram Help Center. See https://about.instagram.com/blog/announcements/safer-internet-day-2020/
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Instagram has implemented a built-in reporting option, so users may report content that violates the Community Guidelines. Instagram has a global team that reviews those reports and removes content that violates its guidelines. Instagram discloses that it may work with law enforcement, including when it believes that there is risk of physical harm or threat to public safety. Also, since Instagram is part of the 'Facebook Companies', it may use the same methods as Facebook to identify and remove TVEC. Indeed, after the Christchurch shootings, a post by Facebook's COO Sheryl Sandberg titled 'By working together, we can win against hate' was published on Instagram's info page (Huang, 2018[96]). The post explained the technology used by Facebook to combat TVEC. This suggests that both platforms use the same technology. The marginal economic costs of using automated tools to identify TVEC are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. Instagram is not a member of the GIFCT, but does participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of ToS or Community Guidelines/Standards	Instagram can remove any content or information users share on the platform if Instagram believes that it violates its ToS and other policies (including the Instagram Community Guidelines). Instagram can also refuse to provide or can stop providing all or part of its service to a user (including terminating or disabling their account) immediately if the user clearly, seriously or repeatedly violates Instagram's ToS and other policies (including the Instagram Community Guidelines). Recently, Instagram announced an update of its account disable policy, explaining that in addition to removing accounts with a certain percentage of violating content (which is undisclosed), it will also remove accounts with a certain number of violations within a window of time (also undisclosed) (Instagram, 2019[97]).
7. Does the service issue transparency reports (TRs) on TVEC?	Yes. Facebook's November 2019 TR included information from Instagram on four areas: "child nudity and sexual exploitation," "regulated goods," "suicide and self-injury," and "terrorist propaganda."
8. What information/fields of data are included in the TRs?	The topic of "terrorist propaganda" contains three fields of information: - *Prevalence (How prevalent were terrorist propaganda, and violence and graphic content, violations on Instagram?)* The prevalence metric is the percentage of

	all views that were of content that violated certain of Instagram's community standards. For example, Facebook/Instagram estimates that less than 0.04% of views were of content that violated its standards for terrorist propaganda in Q1 2019. In other words, fewer than 4 of every 10,000 views on Facebook/Instagram contained what the company deemed to be terrorist propaganda. (The figures refer to final determinations, not content that was initially flagged as a possible violation but may have been subsequently determined to be permissible.) - *Content actioned (How much content did Instagram take action on?)* Taking action may include removing a piece of content from Instagram, covering photos or videos that may be disturbing to some audiences with a warning, or disabling accounts. Content actioned is the total number of pieces of content that Instagram took action on during a given reporting period because it violated its community standards. - *Proactive rate (Of the violating content actioned, how much did Instagram find* before *users reported it?)* This metric shows the percentage of content actioned for violating Instagram's policies that Instagram found and flagged before users reported it. It counts detections made by both Instagram's AI tools and human reviewers.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	- *Prevalence.* The prevalence metric is the estimated number of views of violating content, divided by the estimated number of total content views on Instagram, per reporting period. For example, if the prevalence of terrorist propaganda is 0.18% to 0.20%, that means of every 10,000 content views, 18 to 20 on average were of content that violated Facebook's standards for terrorist propaganda. The prevalence metric provides an indication of how often prohibited content is seen, rather than the total amount of such content published. Prevalence is estimated based on samples of content across different areas of Instagram. For terrorist propaganda violations, in particular, Instagram only estimates the upper limit, which means that Instagram is 'confident that the prevalence of violating views is below that limit.' (Facebook, n.d.[71]) - *Content actioned.* Content actioned is the total number of pieces of content that Instagram took action on during a given reporting period because it violated its content policies. Instagram does not count those scenarios where it escalates content to law enforcement. This metric includes both content Instagram actioned after someone reported it and content that Instagram found proactively. - *Proactive rate.* This metric is calculated as: the

	number of pieces of content actioned that Instagram found and flagged before users reported them, divided by the total number of pieces of content actioned.
10. Frequency/timing with which TRs are issued	Instagram TRs are issued jointly with Facebook's and follow the same reporting schedule.
11. Has this service been used to post TVEC?	Yes. The media has coverfed many examples, (Carmen, 2015[98]) (Hymas, 2019[99]) (Cox, 2019[100]).

8. QQ

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no definition. However, in its ToS, QQ prohibits its users from submitting, uploading, transmitting or displaying any content which in fact or in QQ's reasonable opinion: • breaches any laws or regulations (or may result in a breach of any laws or regulations); • creates a risk of loss or damage to any person; • harms or exploits any person (whether adult or minor) in any way, including via bullying, harassment or threats of violence; and • is hateful, harassing, abusive, racially or ethnically offensive, defamatory, humiliating to other people (publicly or otherwise), threatening, profane or otherwise objectionable.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.tencent.com/en-us/zc/termsofservice.shtml and https://www.tencent.com/en-us/zc/acceptableusepolicy.shtml[4]
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	QQ broadly states that it may review (but make no commitment to review) content (including any content posted by users) or third party services made available through QQ to determine whether or not they comply with QQ's policies, applicable laws and regulations or are otherwise objectionable, and QQ reserves the right to block or remove content for any reason, as required by applicable laws and regulations.

4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	QQ provides no information in this regard. QQ is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	QQ may suspend or terminate access to QQ if it reasonably believes that a user has breached QQ's ToS, their use of QQ creates risk for QQ or other QQ users, the suspension or termination is required by applicable laws, or at QQ's sole and absolute discretion.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

9. Youku Tudou

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, in its ToS, Youku Tudou prohibits content that incites ethnic hatred, ethnic discrimination and/or undermines ethnic unity, as well as content that induces the commission of crimes, glorifies violence, or engages in terrorist activities.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at http://mapp.youku.com/service/agreement-eng

3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Youku Tudou broadly states that it 'manages' the information users upload, release or transmit on the platform, and takes measures such as suspending transmissions, removing uploaded content to prevent further dissemination, saving records and reporting to competent authorities in the event that information uploaded is banned by applicable laws and regulations or constitutes a breach of the ToS.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Youku Tudou provides no information in this regard. Youku Tudou is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Breaches of Youku Tudou's ToS may lead to the removal of content, the blocking of content and information, the suspension, termination or cancelation of a user account, or any other measures that may be taken in accordance with the applicable regulations.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

10. QZone

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no definition. However, QQ International's ToS[5] prohibit users from publishing, delivering, transmitting or storing any content that contravenes the law or any content that is inappropriate, insulting, obscene and violent.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://imqq.com/html/FAQ_en/html/Miscellaneous_1.html [6]
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	No procedure is specified.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	QQ International provides no information in this regard. QQ International is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	QQ International states that breach of its ToS entitles them to interrupt the user licence, stop the provision of services, apply use restrictions, reclaim the user's QQ account, carry out legal investigations and other relevant measures, taking into consideration the severity of the user's conduct, without prior notice to the user.

7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

11. TikTok

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, TikTok's Community Guidelines provide that 'dangerous individuals or organisations' cannot use Tiktok to promote terrorism, crime, or other types of behaviour that could cause harm. Terrorists and terrorist organisations are expressly included within that group. TikTok defines 'terrorists and terrorist organisations' as any non-state actors that use premeditated violence or threats of violence to cause harm to non-combatant individuals, in order to intimidate or threaten a population, government, or international organisation in the pursuit of political, religious, ethnic, or ideological objectives. More broadly, TikTok defines 'dangerous individuals and organisations' as those that commit crimes or cause other types of severe harm. The types of groups and crimes include, but are not limited to Hate groups, Violent extremist organizations, Homicide, Human trafficking, Organ trafficking, Arms trafficking, Drug trafficking, Kidnapping, Extortion, Blackmailing, Money laundering, Fraud, Cybercrime. Names, symbols, logos, flags, slogans, uniforms, gestures, portraits, or other objects meant to represent dangerous individuals and/or organisations, or content that praises, glorifies, or supports dangerous individuals and/or organisations is prohibited on TikTok, except for educational, historical, satirical, artistic, and other content that can be clearly identified as counterspeech or aims to raise awareness of the harm caused by dangerous individuals and/or organisations.

2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.tiktok.com/en/terms-of-use#terms-eea, http://support.tiktok.com/en/privacy-safety/community-policy-en and https://www.tiktok.com/community-guidelines?lang=en
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	TikTok broadly states that it may, at any time and without prior notice, remove or disable access to content at its discretion for any reason or no reason. The removal of content may be based on TikTok finding the content objectionable, in violation of its ToS or Community Guidelines, or otherwise harmful to its services or users.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	If a user believes Tiktok has removed their content by mistake, they can appeal this decision.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	TikTok has a reporting mechanism that allow users to report anything that violates TikTok's Community Guidelines. TikTok uses 'automated systems' to alert its staff of problematic content or accounts. Content or accounts flagged as potentially associated with terrorism or extremism are then reviewed by expert members of TikTok's moderation team. Also, Tiktok trains its moderation team in the latest techniques used by terrorists to try to avoid detection, as and when such techniques are discovered. This is in addition to training that all moderators receive in how to spot terrorist content and accounts and distinguish them from other problematic yet allowed content or accounts. The marginal economic costs of using automated tools to identify TVEC are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. TikTok is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violation of the Community Guidelines may result in account termination and/or content removal.

7. Does the service issue transparency reports (TRs) on TVEC?	No. However, TikTok published its first TR in December 2019, disclosing legal requests for user information, government requests for content removal, and copyright content take-down notices for the first half of 2019 (TikTok, 2019[25]).
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

12. Weibo

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, Weibo's ToS prohibit users from uploading, displaying and transmitting any content that is offensive, abusive, intimidating, racially discriminatory, malicious, violent or otherwise illegal.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.weibo.com/signup/v5/protocol
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Weibo broadly states that its operators have the right to review, supervise and process the behaviour and information of Weibo users, including but not limited to user information (account information, personal information, etc.), content data (location, text, pictures, audio, video, trademarks, patents, publications, etc.), and user behaviour (relationships, comments, private letters, participation topics, participation activities, marketing information, complaints, etc.).
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.

4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Weibo has a reporting mechanism that allow users to report unlawful or objectionable content. These reports are verified and processed by moderators. The marginal economic costs of using human moderators to identify objectionable content are probably relatively high. Weibo is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violation of the ToS entitles Weibo to discontinue or terminate the provision of its services.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. The Christchurch shooting was posted on Weibo (Kenny, 2019[95]).

13. iQIYI

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no definition. However, iQIYI's ToS prohibit the promotion of terrorism, extremism (not specifically violent extremism), hatred, ethnic discrimination and dissemination of violence.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.iqiyi.com/user/register/protocol.html

3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	iQIYI broadly state that it reserves the right to cancel users' access to its products and services, or their ability to create, upload, publish and disseminate content, without prior notice.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	iQIYI provides no information in this regard. iQIYI is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	iQIYI notes that violations of its ToS give iQIYI the right to suspend or cancel the infringer's account, and report certain violations to the authorities, where appropriate.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

14. Reddit

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, Reddit's Content Policy prohibits content that encourages, glorifies, incites, or calls for violence or physical harm against an individual or a group of people.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.redditinc.com/policies/user-agreement and https://www.redditinc.com/policies/content-policy It is important to note that Reddit employs a layered moderation system. While the Content Policy above governs all content on Reddit, the site itself consists of thousands of individual communities that are created and moderated by users themselves, on a volunteer basis. These moderators set their own community rules, unique to each specific community depending on its topic, in addition to the sitewide Content Policy. These rules are clearly marked in the sidebars of each individual community.
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	Yes. Available at https://www.redditinc.com/policies/broadcasting-content-policy. In addition to the normal Content Policy, livestreamed content on Reddit is also subject to additional rules: **No NSFW Content** Broadcasts on Reddit may not include NSFW ("Not Safe for Work") content. As noted in the Content Policy, this means content that contains nudity, pornography or sexually suggestive content, or graphic violence, which a reasonable viewer may not want to be seen accessing in a public or formal setting such as a workplace. **No Illegal or Dangerous Behavior** Broadcasts may not contain activities that are illegal, or that pose unreasonable risk of bodily harm to the stream subject or bystanders. **No Quarantine-Eligible Content** Broadcasts on Reddit may not include content that would otherwise trigger a Quarantine. As noted in the Content Policy, this means content that average 'redditors' may find highly offensive or upsetting, or which promotes hoaxes.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other	At the sitewide level, Reddit administrators (paid Reddit employees) have a variety of different methods to enforce their rules, including: • Asking the user nicely to 'knock it off' • Asking the user less nicely

enforcement decisions and appeal processes against them?	• Temporary or permanent suspension of accounts • Removal of privileges from, or adding restrictions to, accounts • Adding restrictions to Reddit communities, such as adding "Not safe for work" tags or quarantining (see below) • Removal of content • Banning of Reddit communities Additionally, volunteer user-moderators also have a number of enforcement methods that they use to enforce rules at the community-specific level. This may include banning the user from that community (either permanently or temporarily), or removing their posts from the community. These actions happen independently of Reddit administrators. Quarantining (Reddit Inc., n.d.[101]) is a measure applied to communities (essentially, groups that share common interests) that average users may find offensive or upsetting, or that are dedicated to promoting hoaxes that warrant additional scrutiny. Its purpose is to prevent the quarantined community's content from being accidentally viewed by those who do not knowingly wish to do so, or viewed without appropriate context. Quarantined communities display a warning that requires users to explicitly opt-in to viewing the content. They generate no revenue, do not appear in non-subscription-based feeds (e.g. Popular), and are not included in search or recommendations. Reddit may also enforce a number of additional product restrictions that exist currently or as it may develop in the future (e.g. removing custom styling tools).
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	To be removed from quarantine, community moderators (see section 5 below) may file an appeal. The appeal should include a detailed account of changes to community moderation practices (appropriate changes may vary from community to community and could include techniques such as adding more moderators, creating new rules, employing more aggressive auto-moderation tools, adjusting community styling, etc.). The appeal should also offer evidence of sustained, consistent enforcement of these changes over a period of at least one month, demonstrating meaningful reform of the community. Reddit may, in its sole discretion, delete or remove content at any time and for any reason, including for a violation of its ToS or Content Policy, or if the content otherwise creates liability for them. Whether applied against an individual account or an entire community, actions taken by Reddit in response to Content Policy violations may be appealed. Reddit employees evaluate the appeals.
5. Means of identifying TVEC (for example, monitoring algorithms, user	Reddit relies on a regime of volunteer user-moderators. Moderating a Reddit community is an unofficial, unpaid position. Community creators are automatically that

generated, human (staff) reviewers, hash-sharing/URL sharing database)	community's first moderators, and they may appoint other users to be moderators to help them as well. Reddit reserves the right to revoke or limit a user's ability to moderate at any time and for any reason or no reason, including for a breach of its ToS. Moderators must follow the Moderator Guidelines (Reddit Inc., 2017[102]), and when they receive reports related to their community, they must take action to moderate by removing content and/or escalating to Reddit administrators for review. Moderators may create and enforce rules for the communities they moderate, provided that such rules do not conflict with Reddit's ToS and other policies. Moderators can set up AutoModerator, which is a site-wide moderation tool assisting the moderation of communities. It enables moderators to carry out certain tasks automatically, such as replying to posts with helpful comments like pointing users to subreddit rules and removing or tagging posts by domain or keyword (Reddit Inc., n.d.[103]). In addition, specially trained Reddit employees are in charge of enforcing Reddit's Content Policy at the sitewide level. Finally, individual Reddit users themselves also participate in flagging and ranking questionable content. Users may report content to either community moderators or Reddit employees. Each user may also downvote a piece of content. Sufficient numbers of downvotes result in the downranking or hiding of the content. The marginal economic costs of using automated tools to identify objectionable content are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. Reddit incurs no costs with regard to user moderators. Reddit is not a member of the GIFCT, but does participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	A violation of Reddit's ToS or Content Policy may lead to the removal of the violating content and/or temporary suspension or permanent termination of the infringer's account (depending on the severity of the incident), status as a moderator, or ability to access or use Reddit's services. Moderators must also follow the Moderator Guidelines, and failing to comply with them also has consequences, including, for example, loss of certain functionalities or moderator privileges. Finally, in the case of communities, if the community itself is not in compliance with Reddit's Content Policy or Moderator Guidelines, the community may be quarantined or banned, depending on the scale or seriousness of the violations.

7. Does the service issue transparency reports (TRs) on TVEC?	Not specifically. However, Reddit does issue Transparency reports that include a section on content removals based on violation of individual community rules or Reddit's Content Policy, which includes the posting of violent content. In its last report (Reddit Inc., 2018[104]), Reddit explained that the vast majority of content removals on Reddit are executed within individual subreddits (communities) by subreddit moderators. These removals are largely based on individual subreddit rules that are unique to each community and set by the moderators and communities themselves. While there may be overlap between enforcement of these rules and Reddit's Content Policy, moderator actions are entirely separate from removals done by Reddit administrators.
8. What information/fields of data are included in the TRs?	The report discloses the number of pieces of content removed by subreddit moderators and by Reddit administrators for violations of the Content Policy; the number of actionable and non-actionable reports for Content Policy violations, and the percentage of Content Policy violations divided by categories of violations (Uncategorised, Impersonation, Personal Information, Minor Sexualisation, Controlled Goods, Involuntary Pornography, Ban Evasion, Harassment and Encouraging Violence or Self-harm).
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not disclosed.
10. Frequency/timing with which TRs are issued	On a yearly basis.
11. Has this service been used to post TVEC?	Yes. The footage of the Christchurch attack was made available in one of Reddit's communities. (Hatmaker, 2019[105]) This led to Reddit administrators banning the entire community in question from the site.

15. Twitter

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition of terrorist or violent extremist *content*, but there is a specific policy on Terrorism and Violent Extremism that includes information on what Twitter considers to be a terrorist or violent extremist organisation, along with examples of content that violates the company's Terrorism and Violent Extremism Policy. In the 'Safety' section of the 'Twitter Rules', terrorism and violent extremism are explicitly forbidden. Also, Twitter has a specific policy on Terrorism and Violent Extremism, under which users may not threaten or promote terrorism or violent extremism. Twitter asserts that there is no

	room in Twitter for terrorist organisations or violent extremist groups and individuals who affiliate with and promote their illicit activities. Twitter's assessments in this context are informed by national and international terrorism designations; however, these designations are not specified. Twitter also assesses organisations under its violent extremist group criteria. Organisations that: • identify through their stated purpose, publications, or actions as an extremist group; • have engaged in, or currently engage in, violence and/or the promotion of violence as a means to further their cause; and • target civilians in their acts and/or promotion of violence are deemed to be violent extremist groups. Twitter examines a group's activities both on and off Twitter to determine whether it engages in and/or promotes violence against civilians to advance a political, religious and/or social cause. Twitter provides the following examples of content that violates its Terrorism and Violent Extremism Policy: • engaging in or promoting acts on behalf of a terrorist organisation or violent extremist group; • recruiting for a terrorist organisation or violent extremist group; • providing or distributing services (e.g., financial, media/propaganda) to further a terrorist organisation's or violent extremist group's stated goals; and • using the insignia or symbols of terrorist organisations or violent extremist groups to promote them. In addition, Twitter's Hateful Conduct Policy provide that users may not promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. Accounts whose primary purpose is inciting harm towards others on the basis of these categories are prohibited. Also, users may not use hateful images or symbols in their profile image or profile header, nor may they use usernames, display names, or profile bios to engage in abusive behaviour, such as targeted harassment or expressing hate towards a person, group, or protected category. This policy bans violent threats, wishing, hoping or calling for serious harm on a person or group of people, references to mass murder, violent events, or specific means of violence where protected groups have been the primary targets or victims, and inciting fear about a protected category. Lastly, Twitter's Glorification of Violence Policy prohibits the

	glorification of violence, especially violent events where people were targeted on the basis of their protected characteristics (including: race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease), as this could incite or lead to further violence motivated by hatred and intolerance. Under this policy, users cannot glorify, celebrate, praise or condone violent crimes, violent events where people were targeted because of their membership in a protected group, or the perpetrators of such acts. Glorification is defined to include praising, celebrating, or condoning statements, such as "I'm glad this happened", "This person is my hero", "I wish more people did things like this", or "I hope this inspires others to act". Violations of this policy include, but are not limited to, glorifying, praising, condoning, or celebrating: • violent acts committed by civilians that resulted in death or serious physical injury, e.g., murders, mass shootings; • attacks carried out by terrorist organizations or violent extremist groups; and • violent events that targeted protected groups, e.g., the Holocaust, Rwandan genocide.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at 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 and https://help.twitter.com/en/rules-and-policies/glorification-of-violence
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Twitter has a range of enforcement options that it may exercise when a user violates the Twitter Rules (Twitter, n.d.[106]). a. *Tweet-level enforcement*: applies to content that violates Twitter's policies, but Twitter believes it is in the public interest that such content remain accessible. In this case, the tweet is hidden behind a notice that give users the option to view the content if they wish. These tweets of public interest are not available in the areas Top Tweets, safe search, recommendations via push and notifications tab, email and text recommendations, live event timeline and explore tab. Also, Twitter takes action at the Tweet level to ensure that it is not being overly harsh with an otherwise healthy account that made a mistake and violated its Rules. Possible tweet level measures

	include limiting tweet visibility, requiring tweet removal and hiding a violating tweet while awaiting its removal. b. *Direct message-level enforcement*: In a private direct message conversation, when a participant reports the other person, Twitter will stop the violator from sending messages to the person who reported them. The conversation will also be removed from the reporter's inbox. In a group direct message conversation, the violating direct message may be placed behind an interstitial to ensure no one else in the group can see it again. c. *Account-level enforcement*: applies when Twitter determines that a person has violated the Twitter Rules in a particularly egregious way, or has repeatedly violated them even after receiving notifications from Twitter. This may include: - Requiring media or profile edits: If an account's profile or media content is not compliant with Twitter's policies, Twitter may make it temporarily unavailable and require that the violator edit the media or information in their profile to come into compliance. Twitter also explains which policy their profile or media content has violated. - Placing an account in read-only mode: If it seems like an otherwise healthy account is in the middle of an abusive episode, Twitter might temporarily make their account read-only, limiting their ability to Tweet, Retweet, or Like content until calmer heads prevail. The person can read their timelines and will only be able to send Direct Messages to their followers. When an account is in read-only mode, others will still be able to see and engage with the account. The duration of this enforcement action can range from 12 hours to 7 days, depending on the nature of the violation. - Verifying account ownership: To ensure that violators do not abuse the anonymity Twitter offers and harass others on the platform, Twitter may require the account owner to verify ownership with a phone number or email address. This helps identify violators who are operating multiple accounts for abusive purposes and take action on such accounts. When an account has been locked pending completion of a challenge (such as being required to provide a phone number), it is removed from follower counts,

	Retweets, and likes until a phone number is provided. - Permanent suspension: This is the most severe enforcement action. Permanently suspending an account will remove it from global view, and the violator will not be allowed to create new accounts.
4.1 Notifications of removals or other enforcement decisions	Notifications take place typically when Twitter requests a user to modify their behaviour and be in compliance with Twitter's rules (requiring media or profile edits), or in case of permanent account suspension. When Twitter permanently suspend an account, it notifies people that they have been suspended for abuse violations, and explains which policy or policies they have violated and which content was in violation.
4.2 Appeal processes against removals or other enforcement decisions	Users can appeal permanent suspensions if they believe Twitter made an error. Upon appeal, if it is found that a suspension is valid, Twitter responds to the appeal with information on the policy that the account has violated.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Twitter has three primary ways of detecting content that may violate its rules. 1. User reporting: Twitter encourages its users to report violations of the Twitter Rules. Moderators review the reports and decide whether the content in fact violates Twitter's rules. Twitter have a global team that manages enforcement of the Twitter Rules with 24/7 coverage in every supported language on Twitter. 2. Proactive content-based detections Twitter also uses internal, proprietary tools to detect violations of the Twitter Rules, including the posting of TVEC, based on the content that is being posted, for example known videos created by terrorist organisations. 3. Proactive behaviour-based detections Twitter utilises internal, proprietary tools to detect violations of the Twitter Rules, including the posting of TVEC, based on the behaviour exhibited that can be associated with terrorist organisations. Twitter has spoken of developing its anti-spam technology to proactively detect TVEC activity, given the tactics utilised by some groups is in part reminiscent of spam. The marginal economic costs of using automated tools to identify TVEC are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. Twitter is member of the GIFCT and participates in GIFCT's Hash Sharing Consortium.

6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violations of the Terrorism and Violent Extremism policy lead to the immediate and permanent suspension of the violating account. Violations of the Hateful Conduct Policy lead to different penalties, depending on a number of factors including, but not limited to, the severity of the violation and an individual's previous record of rule violations. For example, Twitter may ask someone to remove the violating content and serve a period of time in read-only mode before they can Tweet again. Subsequent violations will lead to longer read-only periods and may eventually result in permanent account suspension. If an account is engaging primarily in abusive behaviour, or is deemed to have shared a violent threat, Twitter will permanently suspend the account upon initial review. Violations of the Glorification of Violence Policy vary depending on the severity of the violation and the account's previous history of violations. The first time a user violates this policy, Twitter requires the user to remove the content. Twitter also temporarily locks the user out of his or her account. If a user continues to violate this policy after receiving a warning, the account will be permanently suspended.
7. Does the service issue transparency reports (TRs) on TVEC?	Yes. Twitter's Transparency Reports (Twitter, 2019[107]) include a section on Twitter Rules enforcement, which include the policies described in Section 1 above.
8. What information/fields of data are included in the TRs?	Twitter discloses the number of unique accounts reported in the reporting period for possible violations of the Twitter Rules. Twitter discloses the number of accounts on which it took action (i.e. unique accounts actioned) based on six categories of the Twitter Rules: abuse, child sexual exploitation (CSE), hateful conduct, private information, sensitive media, and violent threats. Moreover, Twitter reports the number of unique accounts suspended for violations related to promotion of terrorism, and the percentage thereof that was reviewed by Twitter's internal, proprietary tools.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	"Unique Accounts Reported" reflects the total number of accounts that users reported as potentially violating the Twitter Rules. To provide meaningful metrics, Twitter de-duplicates accounts that were reported multiple times (whether multiple users reported an account for the same potential violation, or whether multiple users reported the same account for different potential violations). For the purposes of these metrics, Twitter similarly de-duplicates reports of specific Tweets. This means that even if Twitter receives reports about multiple Tweets by a single user, it counts these reports towards the "Unique Accounts Reported" metric only once.

	"Unique Accounts Actioned" reflects the total number of accounts that Twitter took some enforcement action on during the reporting period. Action may be any of the enforcement options explained in section 4 above. To provide meaningful metrics, Twitter de-duplicates accounts that were actioned multiple times for the same policy violation. This means that if Twitter took action on a Tweet or account under multiple policies, the account would be counted separately under each policy. However, if Twitter took action on a Tweet or account multiple times under the same policy (for example, Twitter may have placed an account in read-only mode temporarily and then later also required media or profile edits on the basis of the same violation), the account would be counted once under the relevant policy.
10. Frequency/timing with which TRs are issued	On a half-yearly basis.
11. Has this service been used to post TVEC?	Yes. See sections 7-8 above.

16. Douban

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, Douban's ToS prohibit users from uploading, distributing and otherwise using content that contains gratuitous violence or promotes violence, racism, discrimination, bigotry, hatred or physical harm of any kind against any group or individual, or which is otherwise objectionable.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.douban.com/note/732773017/
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Douban broadly states that it reserves the right (but have no obligation) to review any user content in its sole discretion. Douban also informs that it may remove or modify user content at any time for any reason, in its sole discretion, with or without notice to the relevant user.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.

4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	No information is provided. Douban is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violations of the ToS entitle Douban to suspend the violator's rights to use its services or terminate the violator's account.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

17. LinkedIn

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	LinkedIn's Professional Community Policy has sections on "Terrorism", "Violence" and "Harmful content and shocking material" that prohibit TVEC: "Terrorism We don't allow any terrorist organizations or violent extremist groups on our platform. And we don't allow any individuals who affiliate with such organizations or groups to promote their activities. Content that depicts terrorist activity, that is intended to recruit for terrorist organizations, or threatens, promotes, or supports terrorism in any manner is not tolerated on the services. "Violence We don't allow any threat of violence against an individual or a group on our platform. This includes statements of an intent to kill or inflict serious physical harm. We don't allow individuals or groups that engage in or

	promote violence, property damage, or organized criminal activity. You may not use our services to express support for such individuals or groups or to post content or otherwise use the services to incite or glorify violence. “Harmful content and shocking material We don’t allow graphic or other content intended to shock or humiliate others. We don’t allow activities that promote, organize, depict, or facilitate criminal activity. We also don’t allow content depicting or promoting instructional weapon making, drug abuse, and threats of theft. Content or activities that promote or encourage suicide or any type of self-injury, including self-mutilation and eating disorders, is also not allowed. If you see signs that someone may be considering self-harm, please report it.”
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.linkedin.com/help/linkedin/answer/34593 (click "Learn more about being safe")
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	Yes. In addition to having to comply with the ToS and the LinkedIn Professional Community Policies, live streaming is a limited feature on LinkedIn. Any member who wants to use it must submit an application and be reviewed under a specific set of criteria. The application form is available here: https://www.linkedin.com/help/linkedin/ask/lv-app LinkedIn has provided additional best practices and guidelines for live streaming, which are available here: https://www.linkedin.com/help/linkedin/answer/100225?query=linkedin%20live&hcppcid=search
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	LinkedIn encourages users to report content that violates its Professional Community Policy. When a user reports another member's content, that other member is not told who made the report, and the reporting user no longer sees the content or conversation they reported in their feed or messaging inbox. LinkedIn may review the reported content or conversation to take additional measures like warning or suspending the author if the content is in violation of its ToS or policies.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	LinkedIn gives users the ability to appeal decisions to restrict content and accounts, as stated in its Professional Community Policies. The process is further explained here: https://www.linkedin.com/help/linkedin/answer/82934

5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users are able to report content that violates LinkedIn's policies. Moderators review the reports to decide whether to take further actions. LinkedIn's parent company, Microsoft, Inc., states that whenever terrorist content on its hosted consumer services is brought to its attention via its online reporting tool, it removes it (Microsoft, 2016[108]). The marginal economic costs of using human moderators to detect objectionable content are probably relatively high. LinkedIn recently became a member of the GIFCT.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	The posting of content that violates LinkedIn's ToS or other policies may lead to a warning or suspension of the author's account.
7. Does the service issue transparency reports (TRs) on TVEC?	Not specifically. LinkedIn issues bi-annual transparency reports (LinkedIn, n.d.[106]) that contain a section on content removal requests from governments reporting violations of its ToS or local laws, as well as a report on content removal under its Professional Community Policies. TVEC is reported as part of the "violent or graphic" category, which "includes content that threatens or promotes terrorism, violence, or other criminal activity, and content that is extremely violent or intended to shock or humiliate others" and thus is broader than TVEC alone. The latest report is available here: https://about.linkedin.com/transparency/community-report
8. What information/fields of data are included in the TRs?	Total content removed. LinkedIn also reports the total number of content removal requests from governments reporting violations of its ToS or local laws, by country, as well as the percentage of requests on which LinkedIn took action, but there is no specific information on removals of TVEC.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	These are described in the community report: https://about.linkedin.com/transparency/community-report
10. Frequency/timing with which TRs are issued	Every six months.
11. Has this service been used to post TVEC?	Possibly. Research has shown that U.S.-based extremists – though not necessarily violent extremists – have used LinkedIn to promote their agendas (START (National Consortium for the Study of Terrorism and Responses to Terrorism), 2018[109]).

18. Baidu Tieba

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, Baidu Tieba's ToS prohibits content that incites hatred based on nationality, ethnic discrimination, violence, murder and terrorism.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://gsp0.baidu.com/5aAHeD3nKhI2p27j8IqW0jdnxx1xbK/tb/eula.html
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	No procedures are specified.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Baidu Tieba has a reporting mechanism that allow users to report unlawful or objectionable content. These reports are verified and processed by moderators, who ultimately make the decision to keep or remove the content. The marginal economic costs of using human moderators to detect objectionable content are probably relatively high. Baidu Tieba is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or	If it deems that a user has violated its ToS, Baidu Tieba may apply a temporary or permanent ban on the infringer, suspend or delete the infringer's account, or impose any other penalties in accordance with

Community Guidelines/Standards	applicable regulations.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

19. Skype

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Skype's parent company is Microsoft. Microsoft's Services Agreement, which governs Skype, prohibits any activity that is harmful to others, such as posting terrorist or violent extremist content, communicating hate speech or advocating violence against others. Microsoft has stated (Microsoft, 2016[108]) that, for the purposes of its services, terrorist content is material posted by or in support of organizations included on the Consolidated United Nations Security Council Sanctions List (United Nations Security Council, n.d.[110]) that depicts graphic violence, encourages violent action, endorses a terrorist organization or its acts, or encourages people to join such groups. The U.N. Sanctions List includes a list of groups that the U.N. Security Council considers to be terrorist organizations. No definition of violent extremism is provided, but Skype's ToS prohibit users from submitting or publishing any content that is hateful, abusive, illegal, racist, offensive or otherwise objectionable in any way.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Microsoft's Services Agreement is available at: https://www.microsoft.com/en-us/servicesagreement. See also https://www.skype.com/en/legal/ios/tos/#1.

3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Skype specifies a notice and take-down procedure. If Skype receives a notification that any material a user posts, uploads, edits, hosts, shares and/or publishes on Skype (excluding private communications) is inappropriate, infringes any rights of any third party, or if Skype wishes to remove that material or content for any reason whatsoever, Skype reserves the right to automatically remove it for any reason immediately or within such other timescales as may be decided from time to time by Skype in its sole discretion. As described in Microsoft's Services Agreement, "If you violate these Terms, we may stop providing Services to you or we may close your Microsoft account. We may also block delivery of a communication (like email, file sharing or instant message) to or from the Services in an effort to enforce these Terms or we may remove or refuse to publish Your Content for any reason. When investigating alleged violations of these Terms, Microsoft reserves the right to review Your Content in order to resolve the issue."
4.1 Notifications of removals or other enforcement decisions	Notifications are at Microsoft's discretion. Microsoft's Services Agreement states: "When there's something we need to tell you about a Service you use, we'll send you Service notifications. If you gave us your email address or phone number in connection with your Microsoft account, then we may send Service notifications to you via email or via SMS (text message), including to verify your identity before registering your mobile phone number and verifying your purchases. We may also send you Service notifications by other means (for example by in-product messages)."
4.2 Appeal processes against removals or other enforcement decisions	Microsoft's Account suspension appeals form is available here: https://www.microsoft.com/en-us/concern/AccountReinstatement
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Microsoft deploys a variety of scanning technology, artificial intelligence, external partnerships, and human moderation operations solutions to detect and investigate TVEC. Furthermore, users are able to report content that violates Skype's ToS or is otherwise unlawful or objectionable. Moderators review the reports to decide whether further action is warranted. Microsoft states that whenever terrorist content on its hosted consumer services is brought to its attention via its online reporting tool, it removes it (Microsoft, 2016[108]).

	The marginal economic costs of using human moderators to detect objectionable content are probably relatively high. Skype is not a member of the GIFCT and does not participate in GIFCT's Hash Sharing Consortium. Microsoft, however, is a founding member of the GIFCT and participates in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Posting content in violation of Skype's ToS or other policies may lead to the termination or suspension of the infringer's Skype account and use of Skype. See also information in Sections 4 and 4.1 above.
7. Does the service issue transparency reports (TRs) on TVEC	No. Microsoft does issue content removal requests reports (Microsoft, 2019[111]), including requests from governments reporting violations of its ToS or local laws, but there is no specific information on removals of TVEC. Moreover, the reports state that they contain numbers that are aggregated across all Microsoft consumer online services "e.g., Bing, Bing Ads, OneDrive, MSN." Thus the reports seem to include Skype, though it is not expressly mentioned.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Possibly. Research by the Counter Extremism Project has found that a number of individuals have accessed and disseminated official extremist (though the source does not expressly specify violent extremist) propaganda materials on Skype (Counter Terrorism Project, n.d.[112]).

20. Quora

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, in Quora's Be Nice, Be Respectful Policy, under the heading 'Banning users in terrorist groups', Quora states that it will ban and delete all the content of any user who is a confirmed and/or declared member of any group on the U.S. State Department list of Foreign Terrorist Organisations.

2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.quora.com/about/tos, https://www.quora.com/about/acceptable_use and https://www.quora.com/What-is-Quoras-Be-Nice-Be-Respectful-policy/answer/Quora-Official-Account
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Quora states that it has the right but not the obligation to refuse to distribute any content on the Quora platform or to remove content. Violations of Quora's policies may lead to a content warning, and if the violator persists with their conduct, they may be prevented from asking questions, writing answers and making comments (edit-blocked) or they may be banned. (Quora, n.d.[113]) Edit-blocks and bans may be temporary; if a person is banned or edit-blocked, they can come back when they cool off and decide to stop their behaviour. Edit-blocks generally last until the person responds via PM and makes their case to be unblocked.
4.1 Notifications of removals or other enforcement decisions	There are no notifications of content removal, but there are content warnings, as specified above.
4.2 Appeal processes against removals or other enforcement decisions	If a user feels that an edit-block or ban was imposed unfairly, then he or she can appeal Quora's decision.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users are able to report content that they believe violates Quora's policies. Reports are sent to the Quora Moderation team for review. The marginal economic costs of using human moderators to detect objectionable content are probably relatively high. Quora is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Content that violates the Be Nice, Be Respectful policy may be reported to and removed by administrators, and violations of this policy can result in a warning, comment-blocking, an edit-block, or a ban (see section 4 above). Depending on the severity of the Be Nice, Be Respectful violation, a user may be banned immediately (i.e., without waiting for content warnings or edit-blocks). Also, Quora may terminate or suspend a user's Quora account for violating any Quora policy.

7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. Questions about how to join a terrorist organisation have been posted on Quora (Lange, 2017[114]).

21. Snapchat

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, in Snapchat's Community Guidelines, under the heading 'Terrorism', Snap states that terrorist organisations are prohibited from using its platform, and Snapchat has no tolerance for content that advocates or advances terrorism. The term 'terrorist organisations' is not defined. Snap also bans any content that promotes discrimination or violence on the basis of race, ethnicity, national origin, religion, sexual orientation, gender identity, disability or veteran status.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.snap.com/en-GB/terms/#terms-row and https://www.snap.com/en-GB/community-guidelines
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	Not applicable. Snapchat does not support livestreaming.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Snap broadly states that it reserves the right to delete any content (i) which they think violates its ToS or Community Guidelines, or (ii) if doing so is necessary to comply with its legal obligations. Snap notes that they support the Santa Clara Principles on Transparency and Accountability in Content Moderation (Santa Clara University's High Tech Law Institute, n.d.[115]),

	which state that companies should provide notice to users whose content is taken down or whose account is suspended about the reason for the removal or suspension. The Principles also state that companies should provide an opportunity for appeal of content removals and account suspensions, but there are as yet no content removal notifications and appeals against content removal decisions or account suspensions specified in Snapchat's policies.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users are able to report content that violates Snapchat's policies (Snap Inc., n.d.[116]). Snap has a dedicated trust and safety team working on a 24/7 basis. Content that is found in violation of Snapchat's policies is removed. The marginal economic costs of using human moderators to detect objectionable content are probably relatively high. Snapchat is not a member of the GIFCT, but does participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	If a user violates Snapchat's ToS or Community Guidelines, Snapchat may remove the offending content, terminate the offender's account, and notify law enforcement. If a user's account is terminated for violations of Snapchat's policies, the infringer is prohibited from using Snapchat again.
7. Does the service issue transparency reports (TRs) on TVEC?	No. Snapchat does issue transparency reports (Snap Inc., 2019[117]) that contain a section on content removal requests from governments reporting violations of its ToS or Community Guidelines, but there is no specific information on removals of TVEC.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. For example, footage of the terrorist attack in Nice, France in 2016 was disseminated on Snapchat's Live stories and Explorer features (Manileve, 2016[118]).

22. Viber

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, Viber's Public Content Policy provides that overly graphic expressions of violence, in particular where the violence is glorified or encouraged, are not allowed on Viber. This includes extreme depictions or descriptions of violence and credible threats of violence to any individual and/or group. Viber prohibits planning or promoting violent acts that could directly or indirectly cause physical or mental harm to others.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.viber.com/terms/viber-terms-use/ and https://www.viber.com/terms/viber-public-content-policy/
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	Not applicable. Viber does not have a livestreaming feature currently.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Viber states that if a user's Public Account is approved by Viber, the user automatically becomes a Public Account Administrator and Public Chat Administrator. Also, upon creating a Community, the user automatically becomes a "Superadmin" of that Community. Administrators must ensure that all content uploaded and displayed in their Public Account or Community complies with Viber's policies, terms of service and all applicable laws and regulations. Administrators may not engage in or permit third parties to engage in any behaviour that is prohibited under any of them. Viber may remove any or all content if they deem that such content is unauthorized or illegal or violates Viber's Policies.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers,	Users have the option to report content that violates Viber's Content Policy. Viber reviews those reports to determine the most suitable course of action. Administrators have the ability to remove violating content from their Accounts and Communities. It is difficult to determine the extent to which Viber is moderated. Viber's Terms of Use provide that Viber does not undertake to monitor Public Chats

hash-sharing/URL sharing database)	or other Forums, and assumes no liability for the content posted therein. In addition, Viber's core features are encrypted), for which reason moderation of content disseminated through those features is not possible. However, the public features such as communities and public chats are not end to end encrypted, and Viber can, upon reports, review them and if required remove them. The marginal economic costs of using human moderators to detect objectionable content are probably relatively high. User moderators entail no cost for Viber. Viber is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Content that violates Viber's policies or that Viber otherwise finds objectionable is removed. In those cases, Viber may suspend or terminate users' accounts, and block participants of Viber Public Chats.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. ISIS announced (Site Intelligence Group Enterprise, 2018[119]) a Nashir News Agency (the ISIS-linked media dissemination group) account on Viber. (Katz, 2019[120]) Viber closed the account immediately after finding it.

23. Pinterest

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, Pinterest's Community Guidelines prohibit images that show gratuitous violence or glorify violence, as well as content used to threaten or organise violence or support violent organisations. The term 'violent organisations' is not defined. Moreover, Pinterest prohibits anything that presents a real risk of harm to people or property, and making threats, organising violence or encouraging others to be violent is not allowed. Pinterest specifically bans any person or group that is dedicated to causing harm to others. This includes terrorist organisations. The term 'terrorist organisations' is not defined.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://policy.pinterest.com/en-gb/terms-of-service and https://policy.pinterest.com/en-gb/community-guidelines
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	Not applicable. Pinterest does not support live streamed content.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Pinterest broadly states that it reserves the right to remove or modify user content, or change the way it is used in Pinterest, for any reason. This includes user content that is considered to be in violation of Pinterest's policies. Pinterest's Community Guidelines note that Pinterest deletes 'some type of content', whereas 'other stuff' are hidden from public areas on its platform, without further elaboration.
4.1 Notifications of removals or other enforcement decisions	Pinterest notifies users when their content is removed 'in most cases', although it is not explained in which specific places notifications indeed take place.
4.2 Appeal processes against removals or other enforcement decisions	There are no appeal processes against a decision to remove content, but account suspensions can be appealed (Pinterest, n.d.[121]).
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Pinterest has a reporting mechanism that allow users to report content that violates its policies. Pinterest has a team of moderators policing content. Terrorist and violent content is removed when detected. Pinterest informs that they collaborate with industry, government and security experts to identify terrorist groups. The marginal economic costs of using human moderators to detect objectionable content are probably relatively high.

	Pinterest is a member of the GIFCT, but does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	In case of violation of Pinterest's policies, Pinterest may terminate or suspend the violator's access to Pinterest immediately, without notice. Notifications of these actions take place at Pinterest's discretion.
7. Does the service issue transparency reports (TRs) on TVEC?	No. Pinterest does issue transparency reports (Pinterest, 2019[122]) that contain a section on content removal requests from governments and private parties reporting violations of its ToS or local laws, but there is no specific information on removals of TVEC.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

24. Vimeo

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no definition. However, Vimeo prohibits any content that promotes or supports "terror or hate groups"; depicts unlawful acts or extreme violence; and provides instructions on how to assemble explosive/incendiary devices or homemade/improvised firearms. Furthermore, members of a "terror or hate group" cannot create a Vimeo account (Cheah, 2019[123]). The term "terror or hate groups" is not defined.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://vimeo.com/terms and https://vimeo.com/help/guidelines
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.

4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Vimeo states that context is of the essence in the application of its rules and processes. When prohibited content appears in the context of a news story or a narrative device in a dramatic work, Vimeo is likely to leave it up. If, however, the overall driving message of the work is to perpetuate a viewpoint that Vimeo has specifically banned, they will remove it. Vimeo also considers a user's speech outside Vimeo (such as social media platforms, blogs, or anywhere else their personal views are clearly represented) in making calls about intent and good faith (Cheah, 2019[123]). As a rule, Vimeo moderators will remove videos that show people being murdered, tortured, or physically or sexually abused, or display shocking, disgusting, or gruesome images. That said, Vimeo understands that there can be videos that engage with these subjects in a critical, thoughtful way. Videos that report on real-world situations sometimes necessarily contain some graphic or violent scenes. Context is important, and documentary or journalistic videos have greater leeway when it comes to depicting violence or the aftermath of violence. To avoid being removed, videos with these elements may not be sensationalistic, exploitative, or gratuitous. They must also be marked with a "Mature" content rating. Videos that recruit for or propagandise terrorist organisations, regardless of whether they show actual violence, are never allowed (Vimeo, n.d.[124]).
4.1 Notifications of removals or other enforcement decisions	Some content removal decisions are notified, such as removals due to copyright infringement. However, Vimeo does not provide users with notice of video or account removals (or a mechanism for appeal) when the removal involves certain categories of prohibited content, such as suspected child abuse material and terrorist content.
4.2 Appeal processes against removals or other enforcement decisions	Copyright-based removals may be appealed. However, there are no appeal processes against a decision to remove TVEC.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users can report any content that violates Vimeo's guidelines and policies. Vimeo states that it may monitor users' accounts, content, and conduct, regardless of their privacy settings. Vimeo has signed an agreement with Active Fence to help identify TVEC content and expects to implement this partnership in early 2020. The marginal economic costs of using human moderators to detect objectionable content are probably relatively high. Vimeo is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.

6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	In case of violation of Vimeo's policies and ToS, Vimeo may, at its option, suspend, delete, or limit access to the infringer's account or any content within it; and terminate the infringing account.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

25. IMO

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no definition. However, IMO's ToS prohibit use of its services to disseminate any threats of violence.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://imo.im/policies/terms_of_service and https://imo.im/policies/acceptable_use_policy.html
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or	IMO broadly states that it reserves the right to remove, screen, edit, or disable access to any content, without notice to the user owning the content, that IMO considers in its sole discretion to be in violation of its policies or otherwise harmful to the IMO Service.

other enforcement decisions and appeal processes against them?	
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	IMO states that they are 'under no obligation to review' content, but it reserves the right to do so at any time. However, it is unclear what manner(s) of review they would undertake. IMO is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violation of IMO's policies may result in the suspension or termination of the infringer's account.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

26. Telegram

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, Telegram's ToS prohibit the promotion of violence on publicly viewable Telegram channels. Notably, that prohibition does not apply to 'Secret Chats'.

2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://telegram.org/tos
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	No procedures are disclosed. Telegram states that if they receive a court order that confirms a user is a terrorist suspect, they may disclose that user's IP address and phone number to the relevant authorities. Telegram also states that so far, this has never happened (Telegram, n.d.[125]).
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Telegram allows users to report content that violates its policies. Telegram also has a team that polices content on public channels. Since 2016, Telegram operates a channel called 'ISIS Watch', which highlights its efforts to delete public channels and bots that promote terrorist content. The channel claims Telegram has removed over 200,000 ISIS public channels and bots (Telegram, n.d.[126]). The marginal economic costs of using human moderators to identify problematic content are probably relatively high. Telegram is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	No sanctions are specified.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.

9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. Several terrorist attacks have been coordinated on Telegram (Bennett, 2019[127]) (Hayden, 2019[128]) (Bennett, 2019[127]) (Hayden, 2019[128]).

27. LINE

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, LINE's ToS prohibit the posting or transmission of violent content. Also, 'activities that benefit or collaborate with anti-social groups' are not allowed. The term 'anti-social group' is not defined.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://terms.line.me/line_terms/
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	Yes, available at https://terms2.line.me/LINELIVE_ToC_ME1
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	LINE discloses a two-step process to monitor posts on its Timeline, LINE LIVE, LINE Manga, LINE Fortune, LINE Pasha, LINE Step, LINE BLOG, LINE Delima and WizBall: First, user-posted content on supported LINE services is checked by LINE's automatic monitoring system to ensure that it does not contain any prohibited language, break any service rules, or violate LINE's ToS or any relevant laws. If objectionable content is found by the monitoring system, it is immediately suspended after being posted. Next, a monitoring team checks any content the monitoring system cannot classify. The monitoring team compares the content against a set of evaluation criteria and previous examples to make a decision on whether or not the content is permitted. If the monitoring team determines the posted content is in violation of LINE's ToS or any applicable laws, it is suspended (LINE, 2019[129]). LINE is unable to monitor any message a user sends/receives on a regular LINE chat room unless the user sends unencrypted chat data to LINE by using the reporting tool

	(LINE, 2019[129]) .
4.1 Notifications of removals or other enforcement decisions	There are no notifications of content removal.
4.2 Appeal processes against removals or other enforcement decisions	A user may appeal removal decisions through LINE's contact form.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users can report any content that violates LINE's policies. Reports are reviewed by LINE's team and they 'take appropriate action' (LINE, n.d.[130]) if they find any violations of such policies. In addition to responding to the user reports, LINE's monitoring system/team actively review the posted content by users (as described in Section 4 above). The marginal economic costs of using automated tools to identify TVEC are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. LINE is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	LINE may delete content, or suspend or delete a user's account, without prior notice, if they believe that the user is violating or has violated its policies.
7. Does the service issue transparency reports (TRs) on TVEC?	No. However, LINE does issue TRs covering three matters: user information disclosure/deletion requests from law enforcement, actions taken against posts that violate LINE's ToS or applicable laws, and message and call encryption deployment status (LINE, 2019[129]).
8. What information/fields of data are included in the TRs?	In the report on the actions taken against violating posts on LINE services, LINE reports the number of content suspended, and percentages assigned to different categories, including Spam, obscene content, solicitation, unpermitted commercial use of accounts, disturbing and problematic content, promotion of illegal activity, and 'others'. TVEC seems to fall within the 'promotion of illegal activity' category (given the examples in Section 9 below), but this not explicitly stated.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	LINE clarifies that disturbing and problematic content may be 'excessively hateful remarks, photos of dead bodies, click fraud, links to phishing sites, etc.', and promotion of illegal activity may include 'announcements of attacks or bombings, sale of illegal drugs, selling online data (such as accounts, coins, and avatars) for real money, etc.'

10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

28. Ask.fm

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, Ask.fm's Community Guidelines state that terrorist organisations and violent extremist groups that intend to encourage or commit terrorist or violent criminal activity are prohibited from maintaining a presence on Ask.fm to promote any of their campaigns or plans, celebrate their violent acts, fundraise, or recruit young people. The terms 'terrorist organisations' and 'violent extremist groups' are not defined. Additionally, users cannot post content that contains any threat of any kind, including threats of physical violence to themselves or others, or incites others to commit violent acts against themselves or others. No explicit definitions of the words "terrorist", "Terrorism" or "extremism" are provided.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://about.ask.fm/legal/2019-07/en/terms.html and https://about.ask.fm/community-guidelines/
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	Not applicable. Ask.fm does not offer any form of live stream capability.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Ask.fm broadly states that they have the right to monitor users' access to or use of its services for violations of its ToS and to review or edit any content. Ask.fm also states that they can block or disable access to any content that they determine is objectionable or harmful to others, without prior notice.
4.1 Notifications of removals or other enforcement decisions	Content that violates Ask.fm's ToS or Community guidelines is removed, in which case the affected user 'may get a warning'. However, it is not specified when this is the case.
4.2 Appeal processes against removals or other enforcement decisions	Users whose accounts have been banned may appeal this decision.
5. Means of identifying TVEC (for example, monitoring algorithms, user	Users are able to report content that they believe violates Ask.fm's policies.

generated, human (staff) reviewers, hash-sharing/URL sharing database)	Reports are sent to Ask.fm's team for review. Ask.fm asserts that they evaluate all reports. Ask.fm also states that they may access users' content and information when they believe it is reasonably necessary to enforce its ToS and protect the safety of Ask.fm's users or members of the public. The marginal economic costs of using human moderators to identify objectionable content are probably relatively high. Ask.fm is not a member of the GIFCT, but does participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violations of Ask.fm's ToS may lead to the suspension or termination of the infringer's account or access to Ask.fm's services.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. It has been reported, for example, that one Ask.fm account offered advice on how to join ISIS fighters in Iraq, as well as what weapons one could expect to be equipped with on arrival. (Miller, 2014[131])

29. Twitch

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, Twitch's ToS provide that acts and threats of violence will be taken seriously and are considered zero-tolerance violations. All accounts associated with such activities will be indefinitely suspended. This includes, but is not limited to: • Attempts or threats to physically harm or kill others • Use of weapons to physically threaten, intimidate, harm, or kill others. Twitch also prohibits hateful conduct, defined as any content or activity that promotes, encourages, or facilitates violence, among other things, based on race, ethnicity, national origin, religion, sex, gender, gender identity, sexual orientation, age,

	disability, medical condition, physical characteristics, or veteran status.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.twitch.tv/p/en-gb/legal/community-guidelines/, https://www.twitch.tv/p/en-gb/legal/terms-of-service/ and https://help.twitch.tv/s/article/about-account-suspensions-dmca-suspensions-and-chat-bans?language=en_US
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Twitch takes enforcement action against accounts that violate its ToS and/or Community Guidelines. Twitch considers several factors when reviewing reports of violations, including the intent and context, the potential harm to the community, legal obligations and others. Depending on the nature of the violation, Twitch takes a range of actions that vary from issuing a warning, imposing a temporary suspension on the account, and for more serious offenses, an indefinite suspension. A warning is a courtesy notice. Twitch may also remove content associated with the violation. Repeating a violation for which a user has been already warned, or committing a similar violation, will result in a suspension. Temporary suspensions range from 24 hours to longer time periods that can exceed 30 days. If an account is suspended, the user may not access or use Twitch's services, including watching streams, broadcasting, and chatting. After the suspension is complete, the user is able to use Twitch's services again. Twitch keeps a record of past violations, and multiple suspensions over time can lead to an indefinite suspension. For the most serious offenses, Twitch immediately and indefinitely suspends the account with no opportunity to appeal.

4.1 Notifications of removals or other enforcement decisions	There are warnings, depending on the nature of the violation.
4.2 Appeal processes against removals or other enforcement decisions	In cases not resulting in immediate suspension, if a user thinks that he or she did not violate Twitch's Community Guidelines, they may submit an appeal in response to an enforcement decision. In the appeal, the user must include the reason they believe the decision was incorrect. Once the appeal has been reviewed, Twitch notifies the user of the result.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Twitch makes available reporting tools to enable users to report content or behaviour that violates Twitch's Community Guidelines, whether in the live broadcast, within the chat, or associated with a video file. Reports are reviewed by Twitch's Safety team, with reports of extreme violence and terrorist content receiving a priority. A second layer of moderation is made possible via Twitch's suite of tools that enables a channel owner (sometimes referred to as broadcaster) to designate other users as moderators of their channel. By doing so, those users then have the ability to ban bad users, remove messages from chat and take the same actions made available to the channel owner. Third Twitch makes available to channel owners a tool that uses machine learning and natural language processing algorithms to prevent the display of messages within chat until they can be reviewed by a channel moderator before appearing to other viewers in the chat. This is referred to as "AutoMod" (Twitch, n.d.[132]). The marginal economic costs of using automated tools to identify objectionable content are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. Twitch incurs no costs with regard to user moderators. Twitch is owned by Amazon, which joined the GIFCT in September 2019.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violations of Twitch's Community Guidelines may lead to removal of content, a strike on the account, and/or suspension of the account. Serious offences are punished with immediate suspension.
7. Does the service issue transparency reports (TRs) on TVEC?	No. Twitch's parent, Amazon, does issue transparency reports (Amazon, n.d.[133]); however, they have no information on TVEC.
8. What information/fields of data are included in the TRs?	Not applicable.

9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	During a coordinated attack on Twitch's service in May 2019, certain users broadcasted offensive content, including past clips from the Christchurch attack. (Marshall, 2019[134]) More recently, a shooter in Halle, Germany livestreamed his attack on Twitch. (British Broadcasting Corporation (BBC), 2019[135]) The attack was viewed by approximately 2,500 users before Twitch removed the footage of that attack, and it did not reappear on the service.

30. Xigua

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, Xigua's ToS prohibit users from promoting terrorism and extremism.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.ixigua.com/user_agreement/
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	No procedures are specified.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.

5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users can report any type of unlawful activity or content on Xigua. Xigua is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violation of Xigua's ToS may lead to the termination of the infringer's account and access to Xigua's services, without prior notice.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

32. Tumblr

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, Tumblr's ToS state that they do not tolerate content that promotes, encourages, or incites acts of terrorism. That includes content which supports or celebrates terrorist organisations, their leaders, or associated violent activities. The term 'terrorist organisations' is not defined.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.tumblr.com/policy/en/terms-of-service and https://www.tumblr.com/policy/en/community
3. Are there specific provisions applicable to livestreamed content in	No.

the ToS or Community Guidelines/Standards?	
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	If Tumblr concludes that a user is violating its policies, they may send the user a notice via email. If the user cannot explain or correct their behaviour, Tumblr may take action against their account. Tumblr notes that it reserves the right to suspend accounts, or remove content, without notice, for any reason, but particularly to protect its services, infrastructure, users, and community.
4.1 Notifications of removals or other enforcement decisions	There are no notifications of content removal.
4.2 Appeal processes against removals or other enforcement decisions	Users may contact Tumblr support to appeal a content removal decision.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users can report any type of unlawful activity or content on Tumblr. Tumblr states that its trained experts review the reported content and take the 'appropriate action´. Reports do not always result in the content being removed. Sometimes Tumblr's experts determine that the reported content does not violate Tumblr's Community Guidelines. Tumblr does use automated tools to identify potentially TVEC-related content for human review, in addition to user reports. The marginal economic costs of using automated tools to detect objectionable content are probably relatively low (although fixed costs may be substantial), whereas the costs of using human moderators are likely relatively high. Tumblr is not a member of the GIFCT, but does participate in the GIFCT's Hash Sharing Consortium.[7]
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Tumblr may terminate or suspend the infringer's access to or ability to use any and all of Tumblr's services immediately, without prior notice or liability.
7. Does the service issue transparency reports (TRs) on TVEC?	No. Oath, previous controller of Tumblr (Alexander, 2019[136]), does release transparency reports. Up until the year 2018, they included Tumblr. However, the reports are very broad and do not break down the information per company controlled by Oath (for example, government requests for removal of content included both Yahoo and Tumblr). Also, there is no information specific to TVEC (Verizon Media, 2019[137]).
8. What information/fields of data are included in the TRs?	Not applicable.

9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. Tumblr is reportedly fraught with pages promoting Nazism, white supremacy, ethno-nationalism, and far-right terrorism (Barnes, 2019[138]) (Fisher-Birch, 2018[139]).

33. Flickr

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, Flickr's ToS do prohibit posting content related to terrorism.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.flickr.com/help/terms and https://www.flickr.com/help/guidelines
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Whilst Flickr relies on a user moderation regime with regard to nudity and indecency, this system does not apply to TVEC, given that posting of TVEC leads to the deletion of the infringer's account. The criteria for identifying TVEC are not specified, though.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users are able to report any content they consider violates Flickr's Community Guidelines. Flickr's staff review such reports to determine whether there is a violation, and take appropriate action. The marginal economic costs of using human moderators to detect objectionable content are probably relatively high.

	Flickr is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Posting TVEC content leads to the deletion of the relevant user's account. Flickr informs that they may report this conduct to law enforcement.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. On Flickr, a virtual monument was created for foreign *jihadi* fighters killed in Syria, featuring their name, origin, and admiring remarks about their devoutness and combat strength (Weimann, 2014[140]).

34. Huoshan

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, Huoshan's ToS ban any content that promotes terrorism and extremism (not specifically violent extremism).
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.huoshanzhibo.com/agreement/
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other	No procedures are specified. Huoshan does inform that it keeps records of alleged violations of laws and regulations and suspected crimes, and report the same to the relevant competent authorities in accordance with the law, cooperating with any relevant investigations.

enforcement decisions and appeal processes against them?	
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users can report any type of unlawful activity or content on Huoshan. Huoshan's team of moderators reviews these reports and takes action accordingly. In addition, Huoshan has staff allocated to content moderation, and is increasing its efforts to improve its 'auditing standards' (Yoo, 2018[141]). The marginal economic costs of using human moderators to detect objectionable content are probably relatively high. Huoshan is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	If a user violates Huoshan's ToS, Huoshan may delete posts or comments, restrict some or all of the functions of the infringer's account, or terminate access to its services.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

35. VK

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, VK's ToS prohibit users from loading, storing, publishing, disseminating, making available or otherwise using any information that contains extremist materials and that promotes criminal activity or contains advice, instructions or guides for criminal activities. VK follows the legal definition of terrorist content provided for in Russian law.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://vk.com/terms and https://vk.com/licence
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	No specific procedures are disclosed. VK broadly states that it reserves the right, at its own discretion as well as upon receipt of information from other users or third parties, to modify (moderate), block or remove any information published in breach of VK's ToS, or suspend, limit or terminate the infringer's access to all or any sections or services of VK at any time, with or without advance notice. Also, VK reserves the right to remove a user's personal page and/or suspend, limit or terminate the user's access to any of VK's services, if VK believes that the user poses a threat to VK and/or its users.
4.1 Notifications of removals or other enforcement decisions	Content removals are notified to users, even content listed in the Federal List of Extremist Materials of the Ministry of Justice of the Russian Federation.
4.2 Appeal processes against removals or other enforcement decisions	If a user disagrees with content being deleted or blocked, they can contact VK Support.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	VK uses a hybrid method of moderation. VK responds to reports from users, regulatory agencies and other organisations, also conducting internal monitoring through 'automatic search and inappropriate content removal mechanisms'. One example of VK's automated tools is the use of digital fingerprints to quickly locate harmful content. Any person can report illegal, offensive, or misleading content with the help of the Report button. VK's moderation team reacts as quickly as possible to ban violators and block content that violates VK's rules or the applicable laws. Also, VK allows users to create 'Communities' and become

	administrators and moderators of them. According to VK's ToS, Community administrators and moderators bear liability for moderation and blocking of content uploaded to the pages that are under control of their communities. In particular, administrators and moderators must delete any content in breach of VK's ToS or applicable laws. The marginal economic costs of using automated tools to identify objectionable content are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. VK incurs no costs with regard to user moderators. VK is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violations of VK's ToS including when creating and administering a Community entitle VK to remove/delete violating content, temporarily block the infringer's access to VK, exclude the content from search results or terminate the infringer's account.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. ISIS accounts have been found in VK (Lokot, 2014[142]).

35. YY Live

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, YY Live's ToS state that users cannot publish, transmit, disseminate, and store violent content.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://zc.yy.com/license.html

3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	No procedures are specified.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	No information is provided. However, research has shown that YY Live implements keyword censorship and surveillance. (Knockell, 2015[143]) Specifically, to enforce its ToS, YY Live has a team within its data security department that maintains "24-hour surveillance" on content and is supported by a system that periodically "sweeps" the platform for offensive content and "automatically" filters keywords. (Knockell, 2015[143]) The marginal economic costs of using automated tools to identify objectionable content are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. YY Live is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	In case of violation of its ToS, YY Live may restrict or freeze the offender's use of their YY account, and restrict or suspend access to one or more specific products, services or functions (such as live video).
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.

9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

36. Medium

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no definition. However, Medium's ToS provide that Medium does not allow content or actions that threaten, encourage, or incite violence against anyone, directly or indirectly; content that promotes violence or hatred against people based on characteristics like race, ethnicity, national origin, religion, disability, disease, age, sexual orientation, gender, or gender identity; posts or accounts that glorify, celebrate, downplay, or trivialize violence, suffering, abuse, or deaths of individuals or groups; and calls for intolerance, exclusion, or segregation based on protected characteristics. The glorification of groups that do any of the above is also prohibited.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://medium.com/policy/medium-rules-30e5502c4eb4 and https://medium.com/policy/medium-terms-of-service-9db0094a1e0f
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	For all user-reported content, Medium takes into account factors like newsworthiness, the context and nature of the posted information, reasonable likelihood, breadth, and intensity of foreseeable social harm, and applicable laws. In evaluating controversial and extreme content (not specifically violent extremist content) under Medium's Rules, moderators employed by Medium apply a risk analysis that includes, at a minimum, the following questions: - What are the foreseeable negative consequences of the information being propagated by Medium, and shared on other social media networks?

	- How severe might the potential impact be? - What is the likelihood of the negative consequence occurring? - Who will likely be affected as a result? - Is there information from nationally and internationally recognized institutions, (such as the CDC, WHO, and other official bodies) to help us determine if content presents an elevated risk? (Medium, n.d.[144]) Medium provides the following example of content areas with elevated risk, which is therefore more likely to be suspended or subject to reduced distribution: Conspiracy theories that have an associated history of harassment or violent incidents among adherents, or theories that may foreseeably incite or cause harassment, physical harm, or reputational harm. (Medium, n.d.[144])
4.1 Notifications of removals or other enforcement decisions	Upon investigating or disabling content associated with a user's account, Medium notifies the user, unless it believes the account is automated or operating in bad faith, or that notifying the user is likely to cause, maintain or exacerbate harm to someone.
4.2 Appeal processes against removals or other enforcement decisions	If a user believes his or her content or account has been restricted or disabled in error, or believes there is relevant context Medium was not aware of in reaching its determination, the user can file an appeal.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users can flag content or accounts that violate Medium's Rules, or file a report containing a description of the alleged violation. Reported posts and users are reviewed by Medium's Trust & Safety team for Rules violations, after which appropriate actions are taken. The marginal economic costs of using human moderators to identify objectionable content are probably relatively high. Medium is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violations of Medium's Rules may result in warnings, account restrictions, limited distribution of posts and content, suspension of content, and suspension of the violating account. Controversial and extreme content (again, not specifically violent extremist content) is particularly likely to be subject to suspended or limited distribution (Medium, n.d.[144]).
7. Does the service issue transparency reports (TRs) on TVEC?	No. Medium issued a TR in 2015 (Medium, 2015[145]) covering government requests for information or content removal in 2014, but there was no specific information on TVEC.

8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

37. Haokan

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no definition. However, Haokan's ToS prohibit the use of its services to provide any substantial support or resources for terrorist operations.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at http://www.haokan88.live/term_condition.html
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Haokan broadly states that they have the right (but not the obligation) to block or remove any content posted on its services, in its sole discretion, especially when the content violates its ToS.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user	Haokan provides no information in this regard. Haokan is not a member of the GIFCT, and does not participate

generated, human (staff) reviewers, hash-sharing/URL sharing database)	in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Haokan informs that violations of its ToS give Haokan the right to terminate or restrict the access to the infringer's account, and to delete any content violating its ToS, without prior notice.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

38. Odnoklassniki

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition. However, Odnoklassniki's ToS ban any propaganda or advocacy of hatred or supremacy based on social, racial, national or religious aspects; any content containing threats or inciting violence or criminal violations; and the publication of any information of extremist nature. The term 'extremist' is not defined.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://ok.ru/regulations
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there	Odnoklassniki broadly states that they may warn, notify or inform users of non-compliance with its ToS. The instructions provided by Odnoklassniki in these cases are mandatory for users.

notifications of removals and appeal processes against removal decisions?	Also, Odnoklassniki explains that they may delete any content which in its opinion violates and/or may violate the applicable laws, its ToS, or cause harm or potential harm to, or threaten the safety of other users or third parties.
4.1 Notifications of removals	Odnoklassniki notifies users of their violations of its ToS at its discretion.
4.2 Appeal processes against removal decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users may become moderators of Personal Pages of other users, or create Groups and become administrators of them. In these cases, they have the obligation to moderate the content posted on such pages and groups. Users can also become moderators of videos and photos, by downloading the Odnoklassniki Moderator App (Odnoklassniki, n.d.[146]). Users can report content that violates Odnoklassniki's ToS. Odnoklassniki's team reviews such reports and decides what actions to take. The marginal economic costs of using employed human moderators to detect objectionable content are probably relatively high. User moderators entail no cost for Odnoklassniki. Odnoklassniki informs that they do not perform and have no technical capability to perform automatic censorship of information in the publicly accessible sections of its Social Network or in the users' Personal Pages, or censorship of personal messages. Nor do they perform pre-moderation of information and content posted by users. Odnoklassniki is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violation of Odnoklassniki's ToS give Odnoklassniki the right to suspend, restrict, or terminate the infringer user's access to its social network.
7. Does the service issue transparency reports (TRs) on TVEC	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.

10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. TVEC content in support of IS has been found on Odnoklassniki (Powell, 2019[147]).

39. Discord

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, Discord's ToS prohibit the sharing of content that directly threatens someone's physical or financial state, as well as any threatened harm to another person or someone related to another person in any capacity. In addition, Discord's ToS provide that users cannot defame, libel, ridicule, mock, stalk, threaten, harass, intimidate or abuse anyone.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://discordapp.com/terms and https://discordapp.com/guidelines
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Discord explains that sharing content that threatens someone's 'physical or financial state' is 'completely unacceptable' and results in immediate content removal and account deletion.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	Users are able to appeal actions taken against their accounts.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users can report any content that violates Discord's ToS and Guidelines. Discord has stated that, although they do not read users' private messages, they do investigate and take immediate appropriate action against any reported ToS violation by a server (something akin to a group or community under a common theme) or user (Liao, 2018[148]). After the report, Discord's 'Trust and Safety' team acts as detectives, looking through the available evidence and gathering as much information as possible. This investigation centres on the reported messages, but can

	expand if the evidence shows that there is a bigger violation — for example, if the entire server is dedicated to bad behaviour, or if the behaviour appears to extend historically. Discord uses "smart computers" and automation to detect spamming and exploitative content such as revenge porn, deep fakes and content threatening child safety, and implements systems such as PhotoDNA to detect that content. However, it is not clear whether Discord uses these systems to detect TVEC. Discord has received reports of servers (something similar to groups of users gathered under a theme) focused on spreading hate speech, harassing others, and convincing others to follow dangerous ideologies. Discord states that they take these reports seriously and remove servers exhibiting extremist (not specifically violent extremist) behaviour. In addition, they assert that they work with law enforcement agencies, third-parties (such as news outlets and academics), and organisations focused on fighting hate (like the Anti-Defamation League and Southern Poverty Law Center) to make sure Discord is up-to-date and ahead of any potential risks. The marginal economic costs of using automated tools to identify objectionable content are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. Discord is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	If a violation of Discord's Community Guidelines is detected, Discord may take any of the following actions regarding users and/or servers: - Removing the content - Warning users and educating them about their violation - Temporary banning as a "cool-down" period - Permanently banning users from Discord and making it difficult for them to create another account - Removing a server from Discord - Disabling a server's ability to invite new users
7. Does the service issue transparency reports (TRs) on TVEC?	No. However, Discord recently issued its first transparency report of any kind, (Discord, 2019[149]) in which they disclose the number of reports they receive for violations of its Community Guidelines, which may include posting TVEC.
8. What information/fields of data are included in the TRs?	Discord's TR covers the period 1 January to 1 April 2019. It discloses the overall number of reports received, as well at the percentage that fell within the Threatening Behaviour category, which is the closest to TVEC.

	The TR also discloses the percentage of the reports for Threatening Behaviour on which Discord took action, but does not disclose whether that action was content removal, a warning, or account deletion. Importantly, Discord acknowledges that they received reports relating to the live-streamed shootings in Christchurch in the evening of the day of the attack. They state that they focused on removing the graphic video as quickly as possible, wherever users may have shared it. Thereafter, Discord saw an increase in reports of 'affiliated content'. They took action to remove users glorifying the attack and impersonating the shooter, as well as on servers dedicated to dissecting the shooter's manifesto, servers in support of the shooter's agenda, and memes concerning the shooting. Discord identifies the approximate number of reports it received about content related to the shootings during the first ten days that followed, as well as the number of account bans and server removals it enforced for related violations.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	No information available.
10. Frequency/timing with which TRs are issued	Undefined.
11. Has this service been used to post TVEC?	Yes. See Section 8 above.

40. Smule

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, Smule's Community Guidelines prohibit any content that incites violence.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.smule.com/en/s/communityguidelines and https://www.smule.com/en/termsofservice
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or	Smule broadly states that it does not pre-screen any user content, but reserves the right to remove or delete any content

Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	in its sole discretion, with or without notice, especially when the content violates its Community Guidelines or ToS. If Smule finds 'objectionable content', it takes appropriate action, including warning the user, suspending or terminating the user's account, removing all of the user's content, and/or reporting the user to law enforcement authorities, either directly or indirectly.
4.1 Notifications of removals or other enforcement decisions	There are notifications in the form of warnings, at Smule's discretion.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users can report any content that violates Smule's ToS and Guidelines. Smule reviews the material flagged by Smule members and may remove it if is deemed inappropriate or unsafe for the Smule community, or if it otherwise violate Smule's Guidelines or ToS. The marginal economic costs of using human moderators to detect objectionable content are probably relatively high. Smule is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	If a user is found in violation of Smule's Guidelines or ToS, Smule may warn the user, remove any offending content, permanently terminate the user's account, notify law enforcement, or take legal action against the infringer.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

41. KaKaoTalk

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, KaKaoTalk's ToS prohibit violent content and behaviour that enables or motivates illegal activities. Also, KaKaoTalk prohibits all forms of discrimination which promotes stereotypes based on region, disability, race, ethnicity, gender, age, job and religion.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.kakao.com/en/terms and https://www.kakao.com/policy/oppolicy?lang=en
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	KakaoTalk broadly states that, in case of violation of its policies or applicable laws, it is able to investigate the breaches, delete the posts in question temporarily or permanently, or restrict all or part of its services temporarily or permanently. Whether the restriction is temporary or permanent depends on the accumulated number of violations; however, any explicit unlawful activities prohibited under applicable laws and regulations lead to permanent restriction, without delay, regardless of the accumulated number of violations.
4.1 Notifications of removals or other enforcement decisions	The enforcement actions above are notified to users via email or other means within the app, at the earliest convenience, except in case of urgent need to protect other users.
4.2 Appeal processes against removals or other enforcement decisions	Users can appeal the actions taken, and KakaoTalk informs appellants of the company's final decision after reviewing the appeal.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users can create a 'story channel', become a master of it and invite managers to work in it. Masters and managers are administrators of story channels and act as moderators. Masters and managers can block and report users and content when they violate KaKaoTalk's policies. In addition, users can report any content that violates KaKaoTalk's policies. KaKaoTalk's team reviews these reports and takes appropriate action. Also, South Korean regulators, such as the National Policy Agency (NPA), the Communications Commissions, and the Korean Communications Standards Commission (KCSC) may request the deletion of any anti-social, violent and illegal information. Moreover, KaKaoTalk can apply restrictions for activities

	prohibited under its policies or in breach of applicable laws and regulations, without any report from users or regulators. Kakao monitors contents in story channels, including blogs and social media, based on keywords concerning TVEC and unlawful content. Kakao TV, Kakao's online video platform, is also subject to content monitoring, including live-streamed content. When problematic content is found on Kakao TV via monitoring, including TVEC, KaKao TV requires the uploader to alter (removing or revising the content) the content. If the content is not revised within 3 days, moderators delete the content and apply a temporary or lifetime ban in proportion to violent nature of the content and the user's aggregate number of violations. However, when it is decided that the content requires imminent action, moderators are authorised to instantly delete the post without delay. The marginal economic costs of using automated tools to identify objectionable content are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. KaKaoTalk incurs no costs with regard to user moderators. KaKaoTalk is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	In case of violations of KaKaoTalk's policies, KaKaoTalk may issue a warning, delete the violating content, and temporarily or permanently restrict its services, depending on the accumulated number of violations. However, any explicit unlawful activities prohibited under the applicable laws and regulations lead to permanent restriction without delay, regardless of the accumulated number of violations.
7. Does the service issue transparency reports (TRs) on TVEC?	No. KaKaoTalk, however, does issue transparency reports (Daum Kakao, n.d.[150]) disclosing the requests of the South Korean government to access user information and remove content, but there is no specific information on TVEC.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

42. DeviantArt

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, DeviantArt's ToS provide that commentaries that are overly aggressive or needlessly abusive are prohibited ('Prohibited Commentaries'). Moreover, users may not use DeviantArt for any unlawful purposes or to upload, post, or otherwise transmit any material that is unlawful, threatening, menacing, harmful or otherwise objectionable.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://about.deviantart.com/policy/service/, https://about.deviantart.com/policy/etiquette/ and https://about.deviantart.com/policy/submission/
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	After prohibited content is reported (a 'deviation'), the 'deviation owner' may receive an anonymous notification asking if the content is, for example, Mature Content, or whatever it was reported as. This gives the owner a chance to address and possibly remedy the situation. If the owner chooses not to take action and the content is not reported again, staff may agree that no deletion or tag is necessary, marking the report invalid. If the number of reports rises, however, it will rise in the staff's queue and they will more quickly take the appropriate action, whether that is adding a tag, deleting the content, or marking the report as invalid. It must be noted that even though a notification is sent to the deviation owner, every report still goes to DeviantArt's staff for final approval. This feature is simply a chance for a user to fix what might be an honest mistake (Kitsune, 2017[151]). Use of any of the communication tools provided by DeviantArt for the purpose of deliberately aggressive or abusive behaviour can result in a disciplinary action (DeviantArt, n.d.[152]). Forum threads that are misplaced, contain inappropriate subject matter, or contain an undesirable number of other violations of DeviantArt's policies are locked and closed to further commentary. As a registered member of DeviantArt, a user is able to participate as an administrator or member of a "Group", which is a set of user pages and applications formed for the purpose of collecting content, discussions and organising members of the site with common interests. Group administrators may determine its own rules and privileges for users who participate in the Group. As a general rule, DeviantArt will not interfere with Groups unless there is a clear violation of its policies. In these cases, DeviantArt can remove a Group and the Group's privileges.

	User accounts found to be demonstrating unacceptable behaviour, by failure to obey DeviantArt's policies or by engaging in abusive or disruptive community activity, can be subjected to a temporary account suspension. (DeviantArt, n.d.[153]) When an account is suspended, visitors to the suspended profile will be greeted by a "Suspended Account" message, which will be displayed instead of the normal profile page for the duration of the suspension. Administrative suspensions can be set for a variable period of time, with typical durations lasting for 24 hours, one (1) week, two (2) weeks, or thirty (30) days (one month). During this time, the profile will lose the ability to make posts, use most elements of the website, or interact with the community in general. The infringer receives notification of the action, which may include a private message or reason concerning why the action was taken, and a timer will be added to the relevant profile page. If the infringer is subject to further disciplinary action, previously recorded suspensions will be factored in. This may lead to a longer suspension or, in the case of repeat offenders, result in any new suspension being escalated to an account termination (DeviantArt, n.d.[154]).
4.1 Notifications of removals or other enforcement decisions	If content is deleted by DeviantArt's staff, the owner gets a notification. Account suspensions are also notified.
4.2 Appeal processes against removals or other enforcement decisions	If the owner believes content is allowed on DeviantArt and the staff made a mistake, the owner can dispute the claim, explaining why. In this case, staff will give it a second consideration. Generally, DeviantArt allows its users to file appeals and make inquiries concerning content removals, violation notices, account suspensions and terminations or other administrative actions. Such appeals, inquiries and questions are reviewed and acted upon by DeviantArt's staff.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Group administrators are content moderators in their Groups. In addition, users can report any content that violates DeviantArt's policies. After a violation is brought to the attention of DeviantArt's staff, they review the report and take appropriate action. DeviantArt states that they have no ability to control the content users may upload, post or otherwise transmit using its service, and do not have any obligation to monitor such content for any purpose. The marginal economic costs of using employed human moderators to detect objectionable content are probably relatively high. User moderators entail no cost for DeviantArt. DeviantArt is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.

6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violations of DeviantArt's policies may lead to a warning, deletion of content, account suspension or termination of the violator's membership, at DeviantArt's sole discretion.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes, Neo-Nazi groups have used DeviantArt to upload propaganda and recruit new members (Hayden, 2019[155]).

43. Meetup

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, according to Meetup's ToS, gratuitously graphic or violent content is prohibited; behaviour that incites violence against individuals or groups of people based on who they are or their beliefs is prohibited; and using Meetup to promote, facilitate, or organise violent, criminal, or non-consensual actions that endanger anyone, physically, mentally or emotionally, is also prohibited. Moreover, 'Groups' (sections within Meetup focused on specific interests or activities) must not contain content or promote events that organise, promote, provide for, distribute services for, or recruit for terrorist organisations; contain content or promote events that could threaten public or personal safety, including advocating for, inciting, or making aspirational statements or threats to commit violence against any group of people, individual person, or specific location, weapons and explosive-making, and calls for violence in response to private or public events.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://help.meetup.com/hc/en-us/articles/360002897532-Usage-and-content-policies-Rules-for-using-Meetup, https://help.meetup.com/hc/en-us/articles/360004285732-Meetup-social-media-community-standards, https://help.meetup.com/hc/en-us/articles/360002897712-Meetup-groups-and-events-policies and https://help.meetup.com/hc/en-us/articles/360027447252-Terms-of-Service

3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Meetup broadly states that violations of its policies and ToS may lead to the modification, suspension or termination of the infringer's account or access to Meetup, and when this happens, they notify the infringer of the reasons for the modification, suspension, or termination.
4.1 Notifications of removals or other enforcement decisions	Enforcement decisions are notified to users.
4.2 Appeal processes against removals or other enforcement decisions	If a user believes the modification, suspension, or termination has occurred in error, he or she can appeal the decision.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Group administrators are content moderators in their Groups, and have the ability to modify, suspend, or terminate users' access to the Groups they moderate. In addition, users can report any content that violates Meetup's policies. Meetup's Trust and Safety team reviews all reports and takes appropriate action. The marginal economic costs of using employed human moderators to detect objectionable content are probably relatively high. User moderators entail no cost for Meetup. Meetup states that they generally do not review content before it is posted (Meetup, 2019[156]). Meetup is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Infringement of Meetup's policies may lead to content deletion, modification, suspension or termination of the infringer's account.
7. Does the service issue transparency reports (TRs) on TVEC?	No. Meetup does issue transparency reports (Meetup, 2017[157]) that disclose government requests for access to users' information and requests for content removal based on Intellectual Property rights infringements, but there is no information on TVEC.
8. What information/fields of data are included in the TRs?	Not applicable.

9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

44. 4chan

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no definition. However, 4chan's ToS prohibit content that violates local or United States laws.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at http://www.4chan.org/rules#global4
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	According to 4chan, threads expire and are pruned by 4chan's software at a relatively fast rate. Since most boards are limited to ten pages, content is usually available for only a few hours or days before it is removed. Usually, missing posts were probably pruned automatically; however, in some cases they may have been removed by a moderator or 'janitor'. Moderators are individuals selected to perform general site maintenance. They may delete posts globally, ban users, close threads and carry out associated actions. Janitors are a class between 'end user' and 'moderator'. They are given access to 4chan's report system and may delete posts on their assigned board(s), as well as submit ban requests. Janitors are selected via an application, orientation, and testing process. Admission to the moderation team is by invitation only. The janitor program is occasionally opened to new applicants. There is no public record of content deletion and because threads are frequently pruned, there is no way of knowing which pieces of content have been removed by the moderation team. In short, there is no way for an end user to

	judge accurately the amount of moderation taking place at any given point in time. The 4chan moderation team reserves the right to block or ban access and remove content for any reason without notice. Users are temporarily blocked from posting when there is a pending ban request placed on their IP address. This block lasts 15 minutes from the time a janitor submits a ban request and is removed immediately if the request is denied by a moderator. If the request is approved, a regular ban is applied.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	Users can appeal bans it they believe an error has been made, by contacting the moderators.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	4chan states that it encourages reporting posts for review (4chan, n.d.[158]). Moderators review the reported content and take appropriate action. The marginal economic costs of using employed human moderators to detect objectionable content are probably relatively high. User moderators entail no cost for 4chan. 4chan is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Breaking 4chan's Rules may result in post deletion, a temporary ban, or in some cases, permanent banishment.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.

11. Has this service been used to post TVEC?	Yes. For example, Neo-Nazi propaganda is common on 4chan (Arthur, 2019[159]).

45. MySpace

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, MySpace's ToS prohibit 'Malicious Bigotry', which is content that actively promotes violence or extreme hatred against individuals or groups on the basis of race, ethnic origin, religion, disability, gender, age, veteran status, or sexual orientation. Encouraging violence or harm to others is strictly prohibited.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://myspace.com/pages/terms#3 and https://help.myspace.com/hc/en-us/articles/202579130-Myspace-Guidelines
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	MySpace broadly states that it reserves the right to investigate and take appropriate action (which may include taking legal action) against anyone who, in Myspace's sole discretion, violates its ToS, including, without limitation, removing the offending content from MySpace, terminating the membership of the violators and/or reporting the violating content or activities to law enforcement authorities. Myspace may seek to gather information from the user who is suspected of violating its ToS and from any other member, and fully cooperates with any law enforcement authorities or court order requesting MySpace to disclose the identity of anyone posting any emails, or publishing or otherwise making available any content that is believed to infringe MySpace's ToS.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user	Users can report content and profiles they think are in violation of MySpace's policies. A MySpace team reviews these reports and may contact the user filing the report to request additional information before making a decision.

generated, human (staff) reviewers, hash-sharing/URL sharing database)	The marginal economic costs of using human moderators to detect objectionable content are probably relatively high. MySpace is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violations of MySpace's ToS can lead to the blocking or deletion of content, and MySpace may consider removing the infringer's profile if they believe the content was posted with the purpose of encouraging violence or harm to others.
7. Does the service issue transparency reports (TRs) on TVEC?	No.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	No.
11. Has this service been used to post TVEC?	Yes. MySpace was once considered a terrorist recruiting ground in the US (Farrell, 2006[160]).

46. Google Drive

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	There is no specific definition of TVEC. However, Google's Abuse Program Policies (Google, n.d.[161]), which apply to Google Drive, have specific provisions on Violence, Hate Speech and Terrorist Content. *Violence*: Users may not threaten to cause serious physical injury or death to a person, or rally support to physically harm others. In cases where there is a serious and imminent physical threat of injury or death, Google may take action on the content. Posting violent or gory content that is primarily intended to be shocking, sensational, or gratuitous is prohibited. If posting graphic content in a news, documentary, scientific, or artistic context, users must provide enough information to help people understand what is going on. In some cases, content may be so violent or shocking that no amount of context will allow that content to remain on Google's platforms. Also, users may not encourage others to commit specific acts of violence. *Hate speech*: Hate speech is not allowed. Hate speech is content that promotes or condones violence against or has the primary purpose of inciting hatred against an individual or group on the basis of their race or ethnic origin, religion, disability, age, nationality, veteran status, sexual orientation, gender, gender identity, or any other characteristic that is associated with systemic discrimination or marginalization.

	Terrorist content: Google does not permit terrorist organizations to use Drive for any purpose, including recruitment. Google also strictly prohibits content related to terrorism, such as content that promotes terrorist acts, incites violence, or celebrates terrorist attacks. The term 'terrorist organizations' is not defined. If users post content related to terrorism for an educational, documentary, scientific, or artistic purpose, they must provide enough information so viewers understand the context.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.google.com/drive/terms-of-service/ and https://support.google.com/docs/answer/148505?visit_id=637064013896463652-1393240150&rd=1
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	No.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	When files are flagged for a violation, the owner of the file may see a flag next to the filename and he or she will not be able to share it. The file will no longer be publicly accessible, even to people who have the link. Users can request that their file be reviewed if they do not think it violates Google's ToS or program policies (Google, n.d.[162]). If a user materially or repeatedly violates Google Drive's ToS or Program Policies, Google may suspend or permanently disable that user's access to Google Drive. Google gives prior notice in such cases. However, Google may suspend or disable a user's access to Google Drive without notice if he or she is using Google Drive in a manner that could cause Google legal liability or disrupt other users' ability to access and use Google Drive.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users can report content that violates Google Drive's ToS and policies. Reports are assessed by Google's staff. Google states that reports do not guarantee removal of the file or any other action on Google's part. This is because content that a user disagrees with or deems inappropriate is not always a violation of Google's ToS or program policies. Google also indicates that they may review users' conduct and content in Google Drive for compliance with the ToS and Program Policies (Google, 2019[163]). Google has reported that files in Google Drive are policed by an algorithm that looks out for abuse of its policies and automatically blocks files that are deemed to violate them. This system

	involves no human review (Titcomb, 2017[164]). The marginal economic costs of using automated tools to identify objectionable content are probably very low (although fixed costs may be substantial), whereas the marginal economic costs of using human moderators to this end are probably relatively high. GoogleDrive is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Abusive material in violation of Google's ToS or other policies entitles Google to: - Remove the file from the account - Restrict sharing of a file - Limit who can view the file - Disable access to one or more Google products - Delete the Google Account (Google, n.d.[165])
7. Does the service issue transparency reports (TRs) on TVEC?	No. Google issues TRs (Google, n.d.[85]) encompassing Google's products and services, including Google Drive. These reports contain a section on government requests to remove content based on violations of local laws or Google's ToS or policies, but there is no TVEC-specific information.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. ISIS content has been found on Google Drive (Katz, 2018[166]).

47. Dropbox

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, Dropbox's Acceptable Use Policy provides that users cannot use Dropbox to publish or share materials that contain extreme acts of violence or terrorist activity, including terrorist propaganda.

2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.dropbox.com/terms and https://www.dropbox.com/terms#acceptable_use
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	Not applicable.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Dropbox states that if a user breaches the ToS or uses Dropbox's services in a manner that would cause a real risk of harm or loss to Dropbox or other users, Dropbox will suspend or terminate the user's access. Dropbox provides reasonable advance notice via the email address associated with the user's account and gives the user an opportunity to export his or her content. If after such notice the user fails to take the steps Dropbox requires, Dropbox will terminate or suspend the user's access to Dropbox's services. Dropbox does not provide advance notice when a user is in material breach of the ToS, when doing so would cause Dropbox legal liability or compromise its ability to provide its services to other users, or when Dropbox is prohibited from doing so by law.
4.1 Notifications of removals or other enforcement decisions	No notifications are specified.
4.2 Appeal processes against removals or other enforcement decisions	No appeal processes are specified.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Users can report content that violates Dropbox's ToS and policies. Dropbox's team reviews these reports, investigates the alleged violation, and takes appropriate action. Dropbox has reported that its staff, on rare occasions, need to access users' file content, particularly to enforce its ToS and policies (Dropbox, n.d.[167]). The marginal economic costs of using human moderators to identify objectionable content are probably relatively high. Dropbox is a member of the GIFCT.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	Violation of Dropbox's ToS or other policies may lead to the suspension or termination of the infringer's account.
7. Does the service issue transparency reports (TRs) on TVEC?	No. Dropbox issues TRs (Dropbox, n.d.[168]) that contain a section on government requests to remove content based on violations of local laws or Dropbox's ToS or policies, but there is no TVEC-specific information.

8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. ISIS content has been found on Dropbox (Bennett, 2019[127]).

48. Microsoft OneDrive

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, Microsoft's Services Agreement (SA), which governs OneDrive, prohibits any activity that is harmful to others, such as posting terrorist or violent extremist content, communicating hate speech or advocating violence against others. Microsoft has stated that for the purposes of its services, they consider terrorist content to be material posted by or in support of organizations included on the Consolidated United Nations Security Council Sanctions List (United Nations Security Council, n.d.[110]) that depicts graphic violence, encourages violent action, endorses a terrorist organization or its acts, or encourages people to join such groups. The U.N. Sanctions List includes a list of groups that the U.N. Security Council considers to be terrorist organizations (Microsoft, 2016[108]).
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://www.microsoft.com/en-us/servicesagreement/
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	Not applicable.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	Microsoft states that it reserves the right to remove or block a user's content from OneDrive at any time if it is brought to its attention that the content may violate applicable law or its SA. When investigating alleged violations of its SA, Microsoft reserves the right to review the user's content in order to resolve the issue. However, Microsoft clarifies that it does not monitor OneDrive. Microsoft follows a "notice-and-takedown" process for

	removal of prohibited content, including terrorist content, which is to say that the "notice" is sent to Microsoft (by a government or a user, for example) and then Microsoft takes down the content. Thus, when the presence of terrorist content on Microsoft's hosted consumer services, including OneDrive, is brought to the company's attention via Microsoft's online reporting tool, Microsoft will remove it (Microsoft, 2016[108]). As described in Microsoft's Services Agreement, "If you violate these Terms, we may stop providing Services to you or we may close your Microsoft account. We may also block delivery of a communication (like email, file sharing or instant message) to or from the Services in an effort to enforce these Terms or we may remove or refuse to publish Your Content for any reason. When investigating alleged violations of these Terms, Microsoft reserves the right to review Your Content in order to resolve the issue."
4.1 Notifications of removals or other enforcement decisions	Notifications are at Microsoft's discretion. Microsoft's Services Agreement states: "When there's something we need to tell you about a Service you use, we'll send you Service notifications. If you gave us your email address or phone number in connection with your Microsoft account, then we may send Service notifications to you via email or via SMS (text message), including to verify your identity before registering your mobile phone number and verifying your purchases. We may also send you Service notifications by other means (for example by in-product messages)."
4.2 Appeal processes against removals or other enforcement decisions	Microsoft's Account suspension appeals form is available here: https://www.microsoft.com/en-us/concern/AccountReinstatement.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Microsoft deploys a variety of scanning technology, artificial intelligence, external partnerships, and human moderation operations solutions to detect and investigate TVEC. Furthermore, users are able to report content that violates Microsoft's policies. Moderators review the reports and decide on the best action to implement. The marginal economic costs of using human moderators to identify objectionable content are probably relatively high. Microsoft is a founding member of the GIFCT and participates in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	If a user posts content that is prohibited or otherwise materially violates the SA, Microsoft may take action against the user, including stopping access to OneDrive, closing the user's Microsoft account immediately, or blocking delivery of a communication (like email, file sharing or instant messaging) to or from the OneDrive. Microsoft may also block or remove infringing content. See also Section 4 above, and this 2016 blog entry:

	"**Observing notice-and-takedown**: We will continue our 'notice-and-takedown' process for removal of prohibited, including terrorist, content. When terrorist content on our hosted consumer services is brought to our attention via our online reporting tool, we will remove it. All reporting of terrorist content – from governments, concerned citizens or other groups – on any Microsoft service should be reported to us via this form." (https://blogs.microsoft.com/on-the-issues/2016/05/20/microsofts-approach-terrorist-content-online/)
7. Does the service issue transparency reports (TRs) on TVEC?	No. Microsoft does issue transparency reports (Microsoft, 2019[111]) that contain a section on content removal requests from governments reporting violations of its ToS or local laws, but there is no specific information on removals of TVEC.
8. What information/fields of data are included in the TRs?	Not applicable.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Yes. ISIS videos have been hosted on OneDrive (Counter Extremism Project, 2018[169]).

49. WordPress.com

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided, though WordPress.com's ToS provide that WordPress.com does not allow websites of terrorist groups recognised by the United States government. The U.S. Department of the Treasury's Office of Foreign Assets Control maintains a list of "Specially Designated Nationals" (US Treasury, 2020[170]), with which WordPress.com is prohibited by law from doing business. WordPress.com does not allow individuals, groups, or entities on that list to use WordPress.com (Word Press, n.d.[171]). Genuine calls to violence are also prohibited. This include the posting of content which threatens, incites, or promotes violence, physical harm, or death, threats targeting individuals or groups, as well as other indiscriminate acts of violence.

2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://en-gb.wordpress.com/tos/ and https://en.support.wordpress.com/user-guidelines/
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	Not applicable.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	WordPress.com has worked in conjunction with experts on online extremism, as well as law enforcement, to develop policies to address extremist (not specifically violent extremist) and terrorist propaganda. WordPress.com suspends websites that call for violence or that are connected to officially banned terrorist groups (per the US Treasury's OFAC list), regardless of content. WordPress.com also implements other measures short of removal—for example, it may flag content and remove a site from the WordPress.com Reader, making the site's content more difficult to find. Flagging a site also removes it from all advertising programs run by WordPress.com. According to WordPress.com, one important way that extremist (again, not specifically violent extremist) sites are brought to its attention is through reports from dedicated government Internet Referral Units (IRUs). These organisations have expertise in online propaganda that private technology companies are not able to develop on their own. They work to identify sites that are being used by known terrorists to spread propaganda or to organise acts of violence. They report terrorist sites to WordPress.com using a dedicated email address that allows WordPress.com to more easily identify reports coming from a trusted source. WordPress.com does not automatically remove websites from WordPress.com. Rather, a human member of its Risk & Safety team reviews each report and makes a decision on whether it violates its policies. One important reason it reviews each report is to guard against the removal of material posted to legitimate sites (news organisations, academic sites) that discuss terrorism or a terrorist group. WordPress.com hosts sites for a number of very large news organisations, news bloggers, academics, and researchers who all publish legitimate reporting on terrorism. In another context, though, some of the materials they publish may qualify as terrorist propaganda, and if so, would be removed under WordPress.com' policies. WordPress.com states that context is very important and they cannot outsource these important decisions affecting legitimate online speech to a robot. Also, since the volume of reports it receives is not high relative to other online platforms, it is able to use more human, versus automated review, when acting on reports (Clicky, 2017[172]).

4.1 Notifications of removals or other enforcement decisions	WordPress.com states that, depending on the scenario, it will email or add a warning notification in the dashboard of a user violating its policies. The notification will contain a link that the user can use to contact WordPress.com regarding the issue. However, those 'scenarios' are not specified (WordPress.com, n.d.[173]).
4.2 Appeal processes against removals or other enforcement decisions	Users can appeal WordPress.com's enforcement actions when the users believe that the actions were taken in error. A real person will review the request and reply with a decision as soon as possible.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	WordPress.com does not pre-screen the content users post. Users are able to report content or sites in violation of WordPress.com's policies. In addition, as noted above, IRUs report terrorist and extremist sites to WordPress.com. WordPress.com evaluates those reports and takes appropriate action. The marginal economic costs of using human moderators to identify objectionable content are probably relatively high. WordPress.com is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.
6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	If WordPress.com finds a site or any of a site's content to be in violation of its policies, WordPress.com will remove the content, disable certain features on the account, and/or suspend the site entirely.
7. Does the service issue transparency reports (TRs) on TVEC?	Yes. Automattic (WordPress.com' parent company) issues TRs that contain a section on reports from IRUs relating to extremist (not specifically violent extremist) content (Automattic, n.d.[174]). The last TR included data from 1 January to 30 June 2019.
8. What information/fields of data are included in the TRs?	- Number of IRU extremist (not specifically violent extremist) content notices - Number of notices for which sites/content were removed as a result - Percentage of notices for which sites/content were removed as a result The figures are broken down by month (January to June) and by reporting country.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	No information available.
10. Frequency/timing with which TRs are issued	On a half-yearly basis. Automattic has issued TRs for the following periods: - 2017: 1 Jul – 31 Dec - 2018: 1 Jan – 30 Jun

	- 2018: 1 Jul – 31 Dec - 2019: 1 Jan – 30 Jun
11. Has this service been used to post TVEC?	Yes. See Section 7 above.

50. Wikipedia

1. How is terrorist and violent extremist content (TVEC) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition is provided. However, the Wikimedia Foundation's ToS, which govern Wikipedia, prohibit harassment, threats, stalking, and vandalism, among other things. The ToS also prohibit using Wikimedia's services in a manner that is inconsistent with applicable law.
2. Manner in which the ToS or Community Guidelines/Standards are communicated	Available at https://foundation.wikimedia.org/wiki/Terms_of_Use/en and https://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines#Enforcement
3. Are there specific provisions applicable to livestreamed content in the ToS or Community Guidelines/Standards?	Not applicable.
4. Policies and procedures to implement and enforce the ToS or Community Guidelines/Standards (removal of content). In particular: are there notifications of removals or other enforcement decisions and appeal processes against them?	The Wikipedia community has the primary role in creating and enforcing its policies. The community is composed of: - *Editors*: volunteers who write and edit the pages of Wikipedia - *Stewards*: volunteer editors tasked with the technical implementation of community consensus, with Checkuser (Wikipedia, 2019[175]) and oversight (Wikipedia, 2020[176]) powers. - *Bureaucrats*: volunteer editors with the technical ability (user rights) to promote other users to administrator or bureaucrat status, remove the admin status of other users, and grant and revoke an account's bot status. - *Administrators*: editors who have been trusted with access to restricted technical features ("tools"). For example, administrators can protect and delete pages, and block other editors (Wikipedia, 2020[177]). Wikipedia's core content policies are: 1. Neutral point of view: All Wikipedia articles and other encyclopaedic content must be written from a neutral point of view, representing significant views fairly, proportionately and without bias. 2. Verifiability: It means that people reading and editing the encyclopaedia can check that information comes from a reliable source.

	3. No original research: Wikipedia does not publish original thought. All material in Wikipedia must be attributable to a reliable, published source (Wikipedia, 2019[178]). Content is deleted by the administrators if it is judged to violate Wikipedia's content or other policies, or the laws of the United States (Wikipedia, 2020[179]). The deletion process encompasses the processes involved in implementing and recording the community's decisions to delete pages and media (Wikipedia, 2020[180]). Normally, a deletion discussion must be held to form a consensus to delete a page. In general, administrators are responsible for closing these discussions, though non-administrators in good standing may close them under specific conditions. However, editors may propose the deletion of a page if they believe that it would be an uncontroversial candidate for deletion. In some circumstances, a page may be speedily deleted if it meets strict criteria set by consensus, which include pages that disparage, threaten, intimidate or harass their subject or some other entity, and serve no other purpose (Wikipedia, 2020[181]). The Wikimedia Foundation states that it rarely intervenes in community decisions about policy and its enforcement. However, when the community requires intervention, or to address an especially problematic user because of significant disturbance or dangerous behaviour, the Wikimedia Foundation may investigate the user's use of the service (a) to determine whether a violation of any policies or laws has occurred, or (b) to comply with any applicable law, legal process, or appropriate governmental request. After the investigation, sanctions may be applied (see Section 6 below).
4.1 Notifications of removals or other enforcement decisions	Not applicable.
4.2 Appeal processes against removals or other enforcement decisions	Not applicable.
5. Means of identifying TVEC (for example, monitoring algorithms, user generated, human (staff) reviewers, hash-sharing/URL sharing database)	Editorial control, and therefore the detection of content that violates Wikipedia's policies, is in the hands of the Wikipedia community. Also, readers (Wikipedia users who do not make contributions) can contact Wikipedia's Volunteer Response Team to report any issue with content on available on Wikipedia. The Wikimedia Foundation states that it does not take an editorial role with respect to its projects, including Wikipedia. This means that it 'generally' does not monitor or edit the content of its projects' websites (Wikimedia Foundation, 2019[182]). The Wikimedia Foundation incurs no costs with regard to Wikipedia community moderators. Wikipedia is not a member of the GIFCT, and does not participate in GIFCT's Hash Sharing Consortium.

6. Sanctions/consequences in case of breaches of the ToS or Community Guidelines/Standards	The Wikipedia community may issue a warning, investigate, delete pages created by, block, and/or ban users who violate the community's policies. The Wikimedia Foundation may refuse, disable, or restrict access to the contribution of any user who violates its ToS, ban a user from editing or contributing or block a user's account or access for actions violating its ToS, and take legal action against users who violate its ToS (including reports to law enforcement authorities).
7. Does the service issue transparency reports (TRs) on TVEC?	No. The Wikimedia Foundation does issue TRs (Wikimedia Foundation, n.d.[183]) covering requests for user data and requests for content alteration and takedown, but there is no section specifically addressing TVEC.
8. What information/fields of data are included in the TRs?	In the section 'Requests for user data', under the heading 'emergency disclosures', the Wikimedia Foundation discloses the number of disclosures of user data in connection with terrorist threats. This does not amount, however, to removals of TVEC.
9. Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
10. Frequency/timing with which TRs are issued	Not applicable.
11. Has this service been used to post TVEC?	Unknown.

Annex C. Definitions

For purposes of this report, the following definitions are provided:

Content: Any type of digital information serving as a medium for TVEC, such as comments, pictures, videos, files, posts, links, chatroom chats, blogs or messages.

Content-Sharing Service: Any online service that enables the transfer, transmission and dissemination of Content, in whatever form, whether one-to-one, one-to-few or one-to-many and irrespective of whether the Content is public-facing, semi-private or private. All of the Services profiled in this Report are Online Content-Sharing Services.

Online Platform: A digital service that facilitates interactions between two or more distinct but interdependent sets of users (whether firms or individuals) who interact through the service via the Internet.

Social Media (or Social Networking) Service: Any online service that allows individuals to build a public or semi-public profile of themselves, upload and access Content shared by other users, interact and establish connections with other users, and express their views and interests.

Terrorist Use of the Internet (TUI): Use of the Internet to promote terrorist aims (for example, using a messaging app to coordinate a terrorist attack). The dissemination of TVEC is a type of TUI whose purpose may be, for instance, to incite violence, radicalise or recruit.

Terrorist and Violent Extremist Content (TVEC): There is no universally accepted definition of terrorism and violent extremism, and congruently, of TVEC. This Report follows the language employed in the Christchurch Call, and uses these terms to refer to the general category of terrorist and violent extremist content on which several Online Content-Sharing Services have policies, make moderation and removal decisions, and in some cases report on in transparency reports.

References

4chan (n.d.), *'Advertise - 4chan'*, http://www.4chan.org/advertise (accessed on 31 August 2019). [63]

4chan (n.d.), *Frequently Asked Questions*, https://www.4channel.org/faq. [158]

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. [136]

Amazon (n.d.), *Amazon.com Help: Law Enforcement Information Requests*, https://www.amazon.com/gp/help/customer/display.html?nodeId=GYSDRGWQ2C2CRYEF. [133]

Apple (n.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. [214]

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. [159]

Artistic license (n.d.), *VK.com TakeDown Process*, https://www.artistic-license.org/takedowns/vk-com-takedown-process/. [207]

Australian Government, F. (2019), *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*, https://www.legislation.gov.au/Details/C2019A00038. [31]

Automattic (n.d.), *Transparency Report*, https://transparency.automattic.com/. [174]

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. [211]

Baidu Inc. (2019), *prnewswire.com*, https://www.prnewswire.com/news-releases/baidu-announces-third-quarter-2019-results-300953076.html. [190]

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* [15]

1934 for the fiscal year ended December 31, 2017.

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/. [138]

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. [127]

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. [196]

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

British Broadcasting Corporation (BBC) (2019), *Germany shooting: 2,200 people watched on Twitch*, https://www.bbc.com/news/technology-49998284. [135]

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/. [123]

Cheng, J. (2014), *South Korea's KakaoTalk Adds 'Secret Mode'*, https://blogs.wsj.com/digits/2014/12/08/south-koreas-kakaotalk-adds-secret-mode/. [208]

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/. [172]

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

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. [184]

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

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. [169]

Counter extremism project (2018), *Extremists & Online Propaganda*, https://www.counterextremism.com/sites/default/files/Extremists%20and%20Online%20Propaga [191]

nda_040918.pdf.

Counter Terrorism Project (n.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 (n.d.), *Transparency Report, Kakao Privacy Policy*, http://privacy.daumkakao.com/en/transparence/report/request. [150]

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, A. (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 (accessed on 5 June 2019). [32]

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

DeviantArt (n.d.), *What happens when my account is banned?*, https://www.deviantartsupport.com/en/article/what-happens-when-my-account-is-banned. [154]

DeviantArt (n.d.), *What is your policy around account suspensions?*, https://www.deviantartsupport.com/en/article/what-is-your-policy-around-account-suspensions. [153]

DeviantArt (n.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. [152]

DeviantArt Media Kit (n.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. [149]

Dropbox (n.d.), *Security - Dropbox, Protecting your files*, https://www.dropbox.com/en_GB/security#files. [209]

Dropbox (n.d.), *Transparency Overview*, https://www.dropbox.com/en_GB/transparency. [168]

Dropbox (n.d.), *Who can see the stuff in my Dropbox account? Dropbox Help*, https://help.dropbox.com/accounts-billing/security/file-access. [167]

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/. [187]

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]

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

European Commission (2018), *Commission Recommendation on measures to effectively tackle illegal content online*, https://ec.europa.eu/digital-single-market/en/news/commission-recommendation-measures-effectively-tackle-illegal-content-online. [35]

European Commission (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 (accessed on 6 June 2019). [34]

European Parliament (2019), *Legislative resolution of 17 April 2019 on the proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online, (COM(2018)0640 – C8-0405/2018 – 2018/0331(COD))*, https://www.europarl.europa.eu/doceo/document/TA-8-2019-0421_EN.html. [36]

Facebook (2020), *Community Standards Enforcement Report: Dangerous Organizations*, https://transparency.facebook.com/community-standards-enforcement#dangerous-organizations (accessed on 5 June 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/. [195]

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 (n.d.), *Community Standards, 2. Dangerous Individuals and Organizations*, https://www.facebook.com/communitystandards/dangerous_individuals_organizations. [69]

Facebook (n.d.), *Community Standards, 2. Dangerous Individuals and Organizations*, [7

https://www.facebook.com/communitystandards/dangerous_individuals_organizations. [70]

Facebook (n.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. [160]

Financial Times (n.d.), *Businesses show no appetite for anti-terror AI tool*, https://www.ft.com/content/fda2d218-56fb-11e9-91f9-b6515a54c5b1. [203]

Financial Times (n.d.), *ISIS videos targeted by UK-funded artificial intelligence software*. [201]

Fisher-Birch, J. (2018), *Terror on Tumblr*, https://www.counterextremism.com/blog/terror-tumblr. [139]

Freedom House (2018), *France Country Report - Freedom on the Net 2018*, https://freedomhouse.org/report/freedom-net/2018/france (accessed on 6 June 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 (accessed on 15 January 2020). [2]

G7 (2019), *G7 Digital Ministers Chair's Summary*, https://www.economie.gouv.fr/files/files/2019/G7/G7Num/Chairs_summary_version_finale_ENG.pdf. [4]

German Federal Office of Justice (2109), *Federal Office of Justice Issues Fine against Facebook*, https://www.bundesjustizamt.de/DE/Presse/Archiv/2019/20190702_EN.html. [42]

Ghoshal, A. (2014), *GitHub, Vimeo and 30 more sites blocked in India over content from ISIS*. [204]

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), *Google Drive Terms of Service*, https://www.google.com/drive/terms-of-service/. [163]

Google (2019), *Transparency report*, https://transparencyreport.google.com/youtube-policy/featured-policies/violent-extremism?hl=en_GB. [188]

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

Google (n.d.), *About - Google Transparency Report*, https://transparencyreport.google.com/about?hl=en_GB. [88]

Google (n.d.), *Abuse program policies and enforcement - Docs Editors Help*, https://support.google.com/docs/answer/148505?visit_id=637064013896463652-1393240150&rd=1. [161]

Google (n.d.), *Google Transparency Report*, https://transparencyreport.google.com/?hl=en_GB. [85]

Google (2010-2019), *Government requests to remove content - Google Transparency Report*, https://transparencyreport.google.com/government-removals/overview?hl=en_GB. [87]

Google (n.d.), *Report a violation - Docs Editors Help*, https://support.google.com/docs/answer/2463296?hl=en&ref_topic=1360897. [165]

Google (n.d.), *Request a review of a violation - Docs Editors Help*, https://support.google.com/docs/answer/2463328?hl=en&ref_topic=1360897. [162]

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

Google, YouTube (n.d.), *Limited features for certain videos - YouTube Help*, https://support.google.com/youtube/answer/7458465. [84]

Google, YouTube (n.d.), *YouTube Community Guidelines enforcement - Violent Extremism*, https://transparencyreport.google.com/youtube-policy/featured-policies/violent-extremism?hl=en_GB&policy_removals=period:Y2019Q2&lu=policy_removals. [86]

Google, Youtube (2020), *Appeal Community Guidelines actions*, https://support.google.com/youtube/answer/185111?hl=en. [76]

Google, Youtube (2020), *Disable or enable Restricted/Safe Mode*, https://support.google.com/youtube/answer/174084?hl=en. [77]

Google, Youtube (2020), *Report inappropriate content*, https://support.google.com/youtube/answer/2802027?hl=en. [78]

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

Google, Youtube (n.d.), *YouTube Community Guidelines enforcement - Hate Speech*, https://transparencyreport.google.com/youtube-policy/featured-policies/hate-speech?hl=en_GB. [82]

Google/Youtube (2020), *Violent Criminal Organizations*, https://support.google.com/youtube/answer/9229472?hl=en&ref_topic=9282436. [75]

Government of France (2018), *Report on online racism and anti-Semitism - Government.fr*, [3

https://www.gouvernement.fr/en/report-on-online-racism-and-anti-semitism (accessed on 6 June 2019). [18]

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. [128]

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. [155]

HM Government (2019), *Online Harms White Paper*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf (accessed on 4 June 2019). [44]

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. [202]

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. [192]

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

ISDGlobal (n.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/. [120]

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/. [166]

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. [151]

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. [143]

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 (accessed on 11 June 2019). [193]

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. [186]

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/ (accessed on 6 June 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. [148]

LINE (2019), *LINE Content Moderation Report*. [129]

LINE (n.d.), *Help Center*, https://help.line.me/line/android/categoryId/20000132/3/pc?lang=en. [130]

LinkedIn (n.d.), *Our Transparency Report*, https://about.linkedin.com/transparency. [197]

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. [118]

Marketing Land (2018), *Quora Introduces Broad Targeting, Says Audience Hits 300 Million Monthly Users*, https://marketingland.com/quora-introduces-broad-targeting-says-audience-hits-300- [51]

million-monthly-users-248261.

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. [134]

Medium (2015), *Medium's Transparency Report (2014)*, https://medium.com/transparency-report/mediums-transparency-report-438fe06936ff. [145]

Medium (n.d.), *Controversial, Suspect and Extreme Content, Medium Help Center*, https://help.medium.com/hc/en-us/articles/360018182453. [144]

Meetup (2019), *Terms of Service*, https://help.meetup.com/hc/en-us/articles/360027447252-Terms-of-Service. [156]

Meetup (2017), *Introducing Meetup's Inaugural Transparency Report*, http://blog.meetup.com/inaugural-transparency-report/. [157]

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 (n.d.), *How OneDrive safeguards your data in the cloud*, https://support.office.com/en-gb/article/how-onedrive-safeguards-your-data-in-the-cloud-23c6ea94-3608-48d7-8bf0-80e142edd1e1. [210]

Miller, J. (2014), *Can Iraqi militants be kept off social media sites?*. [131]

Moign Khawaja, S. (2019), "Disrupting Daesh: Measuring Takedown of Online Terrorist Material and Its Impacts", *Studies in Conflict & Terrorism*, Vol. 42/1-2. [213]

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

New America (n.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 (n.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 (n.d.), *Help Centre*, https://ok.ru/help/54/367. [146]

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]

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. [122]

Pinterest (n.d.), *Account suspension*, https://help.pinterest.com/en/article/account-suspension. [121]

Powell, B. (2019), *Encrypted Extremism - Inside the English-Speaking Islamic State Ecosystem on Telegram*. [147]

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

Reddit Inc. (n.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. [185]

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/. [194]

Rosenthal, M. (ed.) (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. [142]

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#. [199]

Santa Clara University's High Tech Law Institute (n.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. [119]

Snap Inc (n.d.), *Privacy Centre - Our Approach to Privacy*, https://www.snap.com/en-GB/privacy/privacy-by-product. [198]

Snap Inc. (2019), *Privacy Centre, Transparency Report - (1 January - 30 June 2019)*, https://www.snap.com/en-GB/privacy/transparency. [117]

Snap Inc. (n.d.), *Safety Centre, Report a safety concern*, https://www.snap.com/en-GB/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 (n.d.), *ISIS Watch*, https://telegram.me/ISISwatch. [126]

Telegram (n.d.), *Telegram Privacy Policy*, https://telegram.org/privacy. [125]

Tencent (n.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. [205]

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/. [189]

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/. [164]

Twitch (n.d.), *How to Use AutoMod*, https://help.twitch.tv/s/article/how-to-use-automod?language=en_US. [132]

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

Twitter (n.d.), *Our range of enforcement options*, https://help.twitter.com/en/rules-and-policies/enforcement-options. [106]

United Nations Security Council (n.d.), *United Nations Security Council Consolidated List*, https://www.un.org/securitycouncil/content/un-sc-consolidated-list. [110]

US Treasury (2020), *OFFICE OF FOREIGN ASSETS CONTROL - Specially Designated Nationals and Blocked Persons List*, https://www.treasury.gov/ofac/downloads/sdnlist.pdf. [170]

Verizon Media (2019), *Transparency Report*, https://www.verizonmedia.com/transparency/index.html?guce_referrer=aHR0cHM6Ly90cmFuc3BhcmVuY3kub2F0aC5jb20vaW5kZXguaHRtbD9ndWNlX3JlZmVycmVyPWFIUjBjSE02THk5M2QzY3VkSFZ0WW14eUxtTnZiUzgmZ3VjZV9yZWZlcnJlcl9zaWc9QVFBQUFKazduZ3VNWS04dHhtNG9hWFM3TUlkNkxlUWxkMEZ5. [137]

Vimeo (n.d.), *How does Vimeo deal with violent content? - Help Center*, https://vimeo.zendesk.com/hc/en-us/articles/224822427-How-does-Vimeo-deal-with-violent-content-. [124]

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]

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]

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. [140]

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), *Terms of Use - Wikimedia Foundation Governance Wiki*, https://foundation.wikimedia.org/wiki/Terms_of_Use/en. [182]

Wikimedia Foundation (n.d.), *Transparency report*, https://transparency.wikimedia.org/. [183]

Wikipedia (2020), *Administration - Wikipedia*, https://en.wikipedia.org/wiki/Wikipedia:Administration#Human_and_legal_administration. [177]

Wikipedia (2020), *Criteria for speedy deletion - Wikipedia*, https://en.wikipedia.org/wiki/Wikipedia:Criteria_for_speedy_deletion#Procedure_for_administrators. [181]

Wikipedia (2020), *Deletion process - Wikipedia*, https://en.wikipedia.org/wiki/Wikipedia:Deletion_process. [180]

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

Wikipedia (2020), *What Wikipedia is not - Wikipedia*. [179]

Wikipedia (2019), *CheckUser - Wikipedia*, https://en.wikipedia.org/wiki/Wikipedia:CheckUser. [175]

Wikipedia (2019), *Core Content Policies - Wikipedia*, https://en.wikipedia.org/wiki/Wikipedia:Core_content_policies. [178]

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/10.1145/3290605.3300390. [206]

Word Press (n.d.), *Terrorist Activity - Support - Word Press.com*, https://en.support.wordpress.com/terrorist-activity/. [171]

WordPress (n.d.), *Legal Guidelines - Support*, https://en.support.wordpress.com/report-blogs/legal-guidelines/. [212]

WordPress.com (n.d.), *Suspended Content and Sites*, https://en.support.wordpress.com/suspended-blogs/. [173]

Yang, S. (2019), *A War Between Two Chinese Internet Giants: Baidu and ByteDance*, https://equalocean.com/ai/20190505-a-war-between-two-chinese-internet-giants-baidu-and-bytedance. [57]

Yoo, E. (2018), *Huoshan latest video platform to clean up vulgar content*, https://technode.com/2018/04/13/huoshan-clean-up/. [141]

Youku Tudou Inc. (NYSE: YOKU) (n.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] There is no universally accepted definition of terrorism and violent extremism, and congruently, of TVEC. This Report does not presume to provide one and uses these terms following the language employed in the Christchurch Call.

[2] TUI can be seen as the overall problem of the use of the Internet to promote terrorist and violent extremist aims, in whatever form, whereas the dissemination of TVEC is a type of TUI, typically intended to incite violence, radicalise and recruit.

[3] See Annex C for definitions of other terms used in this Report.

[4] See Column 'Type of service' in the tables appearing in Annex A.

[5] "MAU helps to measure an online business's general health and is the basis for calculating other website metrics. MAU is also useful when assessing the efficacy of a business's marketing campaigns and gauging both present and potential customers' experience. Investors in the social media industry pay attention when companies report MAU, as it is a [key performance indicator] that can affect a social media company's stock price." (Investopedia, 2019[192])

[6] Information from media outlets and other publicly available sources was used, however, in Section 10 of each profile (see Annex B), not least because the Services' governing documents rarely list concrete incidents where their technologies are exploited to further terrorist and violent extremist ends. At any rate, when used, these sources of information are duly referenced via footnotes in the relevant profiles.

[7] See Section 1 of the 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 and WordPress.com profiles.

[8] See Section 1 of the Facebook, YouTube, TikTok, Twitter and Google Drive profiles. Arguably, Microsoft (Skype and OneDrive) belongs in this group, as well, though it provides no definition of violent extremism and does not offer any examples.

[9] See Section 1 of the Instagram, Youku Tudou, iQIYI, LinkedIn, Baidu Tieba, Skype, Quora, Snapchat, Pinterest, Ask.fm, Xigua, Tumblr, Flickr, Huoshan, Haokan, Meetup, Dropbox, Microsoft OneDrive and Wordpress.com profiles.

[10] See Section 1 of the WeChat, Instagram, QQ, Youku Tudou, iQIYI, Douban, LinkedIn, Baidu Tieba, Vimeo, Twitch, Medium, Odnoklassniki, KaKaoTalk, Meetup and MySpace profiles.

[11] See Section 1 of the WhatsApp, iMessage/FaceTime, QZone, Weibo, Reddit, Viber, IMO, Telegram, LINE, VK, YY Live, Discord, Smule, DeviantArt, 4chan and Wikipedia profiles.

[12] See Section 7 of the Facebook profile.

[13] See Section 7 of the YouTube profile.

[14] This approach could be criticised on the basis that some governments label rival political groups as terrorists. Quora, conversely, explicitly relies on the U.S. State Department list of Foreign Terrorist Organisations. See Section 1 of the Quora profile. WordPress.com, which relies on the list of 'Special Designated Nationals' of the U.S. Department of the Treasury's Office of Foreign Assets Control, follows a similar approach. See Section 1 of the WordPress.com profile.

[15] See Section 1 of the Skype and OneDrive profiles.

[16] For example, a study conducted based on a dataset comprised of 722 pro-IS accounts and 451 other jihadist accounts on Twitter found that the other jihadist accounts were able to produce 6 times more content and had 13 times more followers than pro-IS accounts on Twitter. Also, whereas 25 percent of pro-IS accounts were suspended within five days of being created, less than 1 percent of the other jihadist accounts were removed within the same timeline. The authors argue that not all jihadists on Twitter are subject to the same levels of disruption as IS. (Moign Khawaja, 2019[213])

[17] For information on the general history of transparency reporting by Internet companies, see https://www.newamerica.org/in-depth/getting-internet-companies-do-right-thing/case-study-3-transparency-reporting/.

[18] See Section 7 of the Facebook, YouTube, Apple, Instagram, TikTok, Reddit, Twitter, LinkedIn, Skype, Snapchat, Pinterest, LINE, Twitch, Tumblr, Medium, Discord, KaKaoTalk, Meetup, Google Drive, Dropbox, OneDrive, WordPress.com and Wikipedia profiles. Note that Facebook Messenger is not counted separately here because Facebook does not issue separate transparency reports for Messenger.

[19] See Section 7-9 of the Facebook, YouTube, Instagram, Twitter and WordPress.com profiles. Note that Discord reported having received reports relating to the live-streamed shooting in Christchurch, but that is the only TVEC-related information they have reported. See Section 8 of the Discord profile. Wikimedia reports the number of "emergency disclosures" of user data based on terrorist threats, but that does not amount to TVEC removals. See Section 8 of the Wikipedia profile. Moreover, LinkedIn reports on violations of its "Violent or graphic" content policy, but in addition to content that threatens or promotes terrorism or violence, and content that is extremely violent, it also includes "other criminal activity" and content that is intended to shock or humiliate others. It is therefore broader than TVEC alone. In addition, please note that Facebook Messenger is not counted separately here because Facebook does not issue separate transparency reports for Messenger.

[20] These include, for example, topics such as climate change, conflict minerals and sexual harassment, exploitation and abuse.

[21] See Section 8 of the WordPress.com profile.

[22] See Section 8 of the Twitter profile.

[23] See Section 8 of the YouTube profile.

[24] See Section 8 of the Facebook profile.

[25] See Section 8 of the Instagram profile.

[26] See Section 8 of the YouTube profile.

[27] See Section 8 of the YouTube profile.

[28] See Section 8 of the Facebook profile.

[29] See Section 8 of the Twitter profile.

[30] See Sections 8-9 of the Facebook profile.

[31] Incidentally, YouTube has reported that 90% of the videos uploaded in September 2018 and removed for Violent Extremism had fewer than 10 views (see https://transparencyreport.google.com/youtube-policy/featured-policies/violent-extremism?hl=en_GB). However, they do not explain the methodology used to arrive at that number.

[32] See Section 8 of the Facebook, YouTube and Twitter profiles.

[33] This is consistent with the Australian Taskforce's proposals. See Section 4 of this Report, Australia.

[34] For example, Twitter's flagging mechanism has been used by white supremacists to shut down accounts of feminists who were using the #solidarityisforwhitewomen hashtag (Crawford, 2014[13]).

[35] See Section 5 of the 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 and Wikipedia profiles.

[36] See Section 4 and 5 of the Reddit, Viber, Twitch, Flickr, VK, Odnoklassniki, KaKaoTalk, DeviantArt, 4chan and Wikipedia profiles.

[37] The expression 'at least' is included because it was not possible to determine, based on some Services' publicly disclosed information, the kind of activities and processes they implement to enforce their ToS and other governing documents. See for example Section 5 of the QQ, Youku Tudou, QZone TikTok, Weibo, iQIYI, Douban, Baidu Tieba, YY Live, Xigua, Huoshan and Haokan profiles.

[38] See Section 5 of the Facebook, YouTube, WhatsApp, Facebook Messenger, WeChat, Instagram (Hash Sharing Consortium member), TikTok, Reddit (Hash Sharing Consortium member), Twitter, LinkedIn (Hash Sharing Consortium member), Skype (indirect membership of GIFCT through Microsoft), Snapchat (Hash Sharing Consortium member), Pinterest (GIFCT member), LINE, Ask.fm (Hash Sharing Consortium member), Twitch (indirect membership of GIFCT through Amazon), VK, YY Live, Google Drive, Dropbox (GIFCT member) and OneDrive (GIFCT member) profiles.

[39] See Section 3 of this Report.

[40] See Section 4.1 of the Facebook, YouTube, Facebook Messenger, Instagram, Reddit, Twitter, Quora, Pinterest, Vimeo, Ask.fm, Twitch, Tumblr, VK, Medium, Odnoklassniki, Smule, KaKaoTalk, DeviantArt, Meetup, Dropbox and Wordpress.com profiles.

[41] See Section 4.2 of the Facebook, YouTube, WhatsApp, Facebook Messenger, Instagram, TikTok, Reddit, Twitter, Quora, Pinterest, Vimeo, LINE, Ask.fm, Twitch, Tumblr, VK, Medium, Discord, KaKaoTalk, DeviantArt, Meetup, 4chan and Wordpress.com profiles.

[42] For example, some services use the expression 'may get a warning' (see Section 4.1 of Ask.fm profile), which suggests that notifications may or may not take place, thus leading to uncertainty. Pinterest follows a similar approach, indicating that users are notified 'in most cases' (see Section 4.1 of Pinterest profile).

Another point in case is Twitch, which notifies users 'depending on the nature of the violation' (see Section 4.1 of the Twitch profile). Smule, Odnoklassniki and Wordpress.com are additional examples (see Section 4.1 of the Smule, Odnoklassniki and Wordpress.com profiles).

[43] See Section 4 and 5 of the WhatsApp, iMessage/FaceTime, WeChat, Instagram, QQ, TikTok, Weibo, iQIYI, Douban, LinkedIn, Quora, Snapchat, Pinterest, IMO, Ask.fm, VK, Haokan, Odnoklassniki, Smule, Meetup, MySpace and OneDrive profiles. Use of the word 'may' or the expression 'reserves the right to review', in particular, are very common.

[44] See Sections 4 and 5 of the WeChat, QQ, Youku Tudou, QZone TikTok, Weibo, iQIYI, Douban, Baidu Tieba, YY Live, Xigua, Huoshan and Haokan profiles.

[45] See Section 7 of the profiles referenced in the preceding endnote.

[46] See statistics at https://www.oberlo.co.uk/blog/tiktok-statistics

[47] In other words, to prevent TUI and the spread of TVEC on their platforms.

[48] The GIFCT has not clarified the meaning of this term.

[49] So far, the United States, United Kingdom, France, Canada, New Zealand, Japan, United Nations Counter-Terrorism Committee Executive Directorate and the European Commission have signed on to the Advisory Committee (GIFCT, 2019[30]). For more information on the GIFCT's governance structure see https://gifct.org/about/

[50] For further information on Australia's abhorrent violent material and ISP blocking schemes, please visit the following references:

- 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

Profiles Notes

[1] This profile is about the Facebook platform itself rather than the entire company, so it does not include Messenger, Instagram or WhatsApp.

[2] The YouTube Trusted Flagger program was developed by YouTube to help provide robust tools for individuals, government agencies, and non-governmental organizations (NGOs) that are particularly effective at notifying YouTube of content that violates their Community Guidelines. https://support.google.com/youtube/answer/7554338?&ref_topic=2803138

[3] See Section 3 of the Report.

[4] It must be noted that these Terms apply only to QQ users anywhere in the world, except if they belong in any of the following categories: (a) a QQ user in China; (b) a citizen of China using QQ anywhere in the world; or (c) a Chinese-incorporated company using QQ anywhere in the world. Users in those categories are governed by the Terms of Service applicable to PRC users, available at https://www.qq.com/contract.shtml

[5] Qzone can be accessed outside China only through QQ International.

[6] These ToS applies to users outside China. QZone users in China are governed by the Terms of Service applicable to PRC users, available at https://www.qq.com/contract.shtml.

[7] Tumblr stated that it participates in the Hash Sharing Consortium; however, as of March 2020, the GIFCT website contains no information about this membership.