CTED TRENDS ALERT

November 2018

MORE SUPPORT NEEDED FOR SMALLER TECHNOLOGY PLATFORMS TO COUNTER TERRORIST CONTENT

CTED
UNITED NATIONS SECURITY COUNCIL
COUNTER-TERRORISM COMMITTEE
EXECUTIVE DIRECTORATE

OVERVIEW

The present *Trends Alert* was prepared by CTED in accordance with Security Council resolution 2395 (2017). This reaffirms the essential role of CTED within the United Nations to identify and assess issues, trends and developments relating to the implementation of Council resolutions 1373 (2001), 1624 (2005) and 2178 (2014) and other relevant resolutions.

CTED *Trends Alerts* are designed to increase awareness, within the Security Council Counter-Terrorism Committee, and among United Nations agencies and policymakers, of emerging trends identified through CTED's engagement with Member States on their implementation of the relevant Council resolutions. The Alerts also include relevant evidence-based research conducted by members of the CTED Global Research Network (GRN)[1] and other researchers.

INTRODUCTION

The collaborative efforts of Member States, international organizations, the private sector and civil society have made it increasingly difficult for terrorist groups - particularly the Islamic State in Iraq and the Levant (ISIL, also known as Da'esh) and its affiliates - to exploit large social media platforms for terrorist purposes. However, measures to counter the overt social media presence of ISIL, and other terrorists and terrorist groups, has caused a shift in their use of the Internet[2] and an increase in their use of smaller, less visible platforms to store and share their material.

TRENDS ALERT

Within the framework of its engagement with Member States, the private sector and civil society, CTED has been alerted to concern that due to this shift, **smaller technology platforms require increased support for their efforts to effectively identify and remove terrorist content, while respecting human rights.**

ISIL has been weakened, the quality of its propaganda has been reduced, and its online dissemination capabilities have been curtailed.[3] However, ISIL retains thousands of supporters who are active online, while ISIL's online messaging remains effective for radicalization, recruitment and instruction purposes.[4]

[1] See September 2018 GRN newsletter for further information.

[2] Shehabat, Ahmad & Mitew, Teodor, Black-boxing the Black Flag: Anonymous Sharing Platforms and ISIS Content Distribution Tactics February 2018.

[3] European Union Terrorism Situation and Trend Report 2018.

[4] The 22nd report of the Analytical Support and Sanctions Monitoring Team on ISIL, Al-Qaida and associated individuals, groups, undertakings and entities.

ISIL and its supporters, and other terrorist groups, including those on the far-right, continue to seek innovative ways to circumvent attempts by large social media platforms to take down or block content, given that these platforms offer terrorist groups the widest possible audience and hence access to potential recruits.

These evolving methods include the use of cloud-based or file-sharing services, URL-shortening services, and the sharing of links to this material using "throwaway" accounts.[5] Simultaneously, terrorists and terrorist groups have continued to migrate to, and between smaller platforms, in search of anonymity and security for their in-group communications, including the sharing of instructional material.

As a result, terrorist content is increasingly located on a diverse range of online platforms. Research by *Tech against Terrorism* (see below) has identified more than 200 platforms actively used by terrorist groups to disseminate content, while one Member State estimates that **between July and December 2017, terrorist material appeared for the first time on almost 150 online services**. Many of these platforms lack the resources required to detect and remove it and are less easily monitored by Government agencies.

AVAILABLE GUIDANCE

Security Council resolutions 2354 (2017), 2395 (2017) and 2396 (2017) stress the need for Member States to act cooperatively to prevent terrorists from exploiting ICT, and to continue voluntary cooperation with the private sector and civil society to develop and implement more effective means to counter use of the Internet for terrorist purposes. This cooperation includes the development of counter-narratives and the implementation of technological solutions, while respecting human rights and fundamental freedoms and ensuring compliance with domestic and international law, as set out in the Comprehensive International Framework to Counter Terrorist Narratives.[6]

The above Council resolutions also recognize the development of the *Tech against Terrorism* initiative and its strategic partner, the Global Internet Forum to Counter Terrorism (GIFCT). *Tech against Terrorism* is a CTED-led initiative that supports technology-industry efforts to tackle terrorist exploitation of the Internet, while respecting human rights. The initiative has numerous partners, including Government, the private sector, civil society, academia, and multi-stakeholder and public/private initiatives.

***Tech against Terrorism* provides extensive guidance for smaller technology companies** on best practices in monitoring online content, countering violent extremism, and facilitating counter-terrorist narratives. It also hosts a Knowledge-Sharing Platform, which offers numerous tools and resources[7] underpinned by six guiding principles, to help

[5] Conway, Maura, Khawaja, Moign, Lakhani, Suraj, Reffin, Jeremy, Robertson, Andrew, & Weir, David, *Disrupting Daesh: Measuring takedown of online terrorist material and its impacts* April 2017.

[6] S/2017/375 (available on the CTC website).

[7] Tech companies can register for the Knowledge-Sharing Platform on the Tech against Terrorism website.

technology companies tackle exploitation of their services while promoting and protecting human rights.

Further guidance includes the joint CTED/ICT for Peace Foundation report on *Private Sector Engagement in Responding to the Use of Internet and ICT for Terrorist Purposes*[8] and the report of the United Nations Office on Drugs and Crime (UNODC) on *The Use of the Internet for Terrorist Purposes*.[9] Guidance is also available from the United Nations Special Rapporteur on Freedom of Opinion and Expression,[10] the Council of Europe,[11] the European Commission,[12] the Global Network Initiative,[13] and the European Union Internet Forum.[14]

CURRENT APPROACHES

Member States' responses to the challenge posed by terrorist use of the Internet have continued to evolve in accordance with the cooperative approach emphasized by Council resolutions 2354 (2017), 2395 (2017) and 2396 (2017).

Through the framework of the GIFCT, major social media companies have expanded their self-regulation methodologies, focusing on responding to breaches of terms-of-service agreements. Using a combination of human and automated moderation techniques, they have developed systems that proactively flag and remove an increasing proportion of terrorist material and the accounts responsible for their transmission. In the first three quarters of 2018, Facebook removed 14.3 million pieces of ISIL- or Al Qaida-related content; 99 per cent of the content removed during Q2 and Q3 was identified by Facebook's internal systems.[15]

The GIFCT fosters collaboration through its shared database of more than 88,000 unique digital hashes that correlate to terrorist content,[16] with a growing number of smaller platforms using and contributing to the database.[17] Some technology firms have taken steps to improve their response to government requests for content removal or data in the context of an investigation, and introduced measures to ensure greater transparency in handling those requests. There have also been initiatives by private sector entities to promote counter-narratives, in partnership with civil society.[18]

[8] Available on the Tech Against Terrorism website.

[9] Available on the UNODC website.

[10] This includes Report to the General Assembly on Artificial Intelligence and Human Rights; The Use of Encryption and Anonymity in Digital Communications; Report to the Human Rights Council on content regulation.

[11] This includes The Rule of Law on the Internet and the Wider Digital World and Filtering, blocking and take-down of illegal content on the Internet.

[12] See ICT Sector Guide on Implementing the UN Guiding Principles on Business and Human Rights.

[13] See Extremist Content and the ICT Sector.

[14] See Tackling Illegal Content Online Towards an enhanced responsibility of online platforms.

[15] Hard questions: what are we doing to stay ahead of terrorists? November 2018

[16] Macdonald, Stuart, How tech companies are successfully disrupting terrorist social media activity June 2018.

[17] Update on the Global Internet Forum to Counter Terrorism.

[18] See e.g., The Redirect Method website.

These global, industry-led initiatives have occurred in parallel to Member States' efforts to criminalize unlawful acts committed by terrorists on the Internet, including the incitement to commit a terrorist act or acts, as required by Security Council resolution 1624 (2005). Some States have gone further by creating new criminal offences relating to accessing terrorist content, or introducing legislation requiring technology companies to remove terrorist content and illegal hate speech from their platforms.

In 2016 for example, the European Commission agreed a voluntary joint code of conduct with Microsoft, Twitter, Facebook and YouTube, on countering illegal hate speech online. In late 2018 however, the Commission proposed a new regulation, stipulating that companies failing to remove illegal terrorist content within one hour of it being flagged by European Union Government authorities could face fines of up to four per cent of their annual global turnover. The regulation also requires service providers to preserve removed material to facilitate appeals against erroneous removal and for the retention of potential evidence. [19]

CHALLENGES

Due to the exponential growth of the Internet and associated technologies, technology platforms have gradually acquired responsibilities that had traditionally belonged to Governments. In the context of terrorist material, those responsibilities include ensuring that their content-moderation policies comply with human rights - including the rights to freedom of opinion and expression, to privacy, to an effective remedy and the principle of non-discrimination - and protect fundamental freedoms.

Significant challenges remain for both large and small technology platforms in this regard. As technology firms increasingly rely on algorithms and artificial intelligence to identify and remove content, concerns have been raised about how they define terrorist content, and about the perceived lack of meaningful human oversight, transparency and accountability.[20] In some instances, legitimate material posted by human rights organizations has been mistakenly blocked or removed. [21]

Despite the significant progress in reducing the availability of terrorist material on the most popular social media platforms, research suggests that it remains accessible. [22] And it is unclear whether the approaches used are fully transferrable to smaller technology platforms, which may be inadequately resourced to monitor material in multiple languages and/or to respond to requests for assistance from Member States across the world. This raises further potential concerns, including:

[19] Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online, September 2018.

[20] UN Human Rights experts meet with Facebook on "overly broad" definitions of terrorist content, September 2018.

[21] Matthews, Kyle & Pogadl, Nicolai, Big Tech is overselling AI as the solution to online extremism, September 2018.

[22] Counter Extremism Project, The eGLYPH Web Crawler: ISIS Content on YouTube, July 2018.

- Difficulties encountered by Governments when requesting digital evidence
- Difficulties encountered by intelligence and law-enforcement agencies when seeking to detect and monitor online terrorist material
- The need to ensure that online counter-narrative campaigns - which have typically used large social media platforms - can effectively reach their target audience by extending their reach to smaller platforms.[23]

Member States are therefore encouraged to continue their cooperative approach to countering terrorist exploitation of the Internet. Guidance developed by *Tech against Terrorism* and the GIFCT for smaller platforms should be part of a holistic, multi-stakeholder approach that focuses on the entire Internet ecosystem and ensures that the necessary support is delivered where and when it is needed.

One potential element of such an approach was trialled in 2018 by one Member State, which developed and released software that could be used by smaller platforms to identify and remove ISIL video content. *Tech against Terrorism* also recommends alternative approaches to online terrorist content required for legitimate purposes such as academic research, including placing flagged content behind login systems.

Member States should also ensure that they strike an appropriate balance between online counter measures and offline prevention efforts[24] and that they and their partners ensure respect for human rights and fundamental freedoms. CTED and *Tech against Terrorism* will continue to engage on these issues in partnership with Member States; relevant United Nations entities; international, regional and subregional organizations; the private sector; civil society and the research community (through the GRN).

[23] Alexander, Audrey, Digital Decay? Tracing change over time among English-language Islamic State Sympathizers on Twitter, October 2017.

[24] CTED-ICT for Peace Private Sector Engagement in Responding to the Use of the Internet and ICT for Terrorist Purposes.