CEP BRIEFING PAPER

The Misuse of Online Platforms by Violent Right-Wing Extremists and Terrorists

Dr. Hans-Jakob Schindler, Alexander Ritzmann and Marco Macori

Counter Extremism Project (CEP) Germany

November 2021

Table of Contents

INTRODUCTION 2

MISUSE OF ONLINE PLATFORMS BY VIOLENT RIGHT-WING EXTREMISTS AND TERRORISTS 3

CHALLENGES AND POTENTIAL COUNTERMEASURES.......... 3

POLICY OPTIONS 5

ANNEX 7

OVERVIEW: EUROPEAN UNION REGULATORY APPROACH 7

EU Terrorism Content Online (TCO) Regulation *7*

EU Digital Services Act (DSA) *7*

OVERVIEW: SELF-REGULATORY APPROACHES.......... 10

EU Internet Forum *10*

EU Internet Forum: EU-wide Crisis Protocol 2019 .... *10*

European Commission and IT Companies: Code of Conduct on illegal online hate speech *11*

Christchurch Call for Action *12*

Tech Against Terrorism *14*

OVERVIEW INDUSTRY LED MECHANISMS.......... 15

Global Internet Forum to Counter Terrorism (GIFCT) *15*

Introduction

In an in-depth study in 2020, the Counter Extremism Project (CEP), commissioned by the Federal Foreign Office of Germany, analysed the transnational connectivity of the violent right-wing extremist (vXRW) and terrorist movement in five European countries and the US.[1] It argued that in particular since 2014 a new leaderless apocalyptic transnational vXRW and terrorist movement emerged, which is responsible for a growing amount of violence in all countries at the centre of the study.

The study also outlined that all governments had developed a variety of countermeasures, ranging from prevention and countering violent extremist (P/CVE) approaches, approaches with a focus on executive and intelligence-led measures to mixed strategies which integrated P/CVE with an increase in executive capacities as well as legal and administrative changes.

However, the study argued that due to the growing transnational connectivity of right-wing extremist and terrorist networks, both offline in physical networking hubs as well as online through specific online ecosystems, transnational measures and mechanisms would be an effective, complementary tool to national strategies and tactics.

From the analysis of the study, five main issue areas emerge, in which further transnational cooperation and coordination could be achieved to mitigate the threat emanating from this movement:

(1) The further development of a common understanding and legal concepts, better capturing the terrorist nature of this developing threat.
(2) Development of a deeper and more nuanced understanding of the various online ecosystems that underpin and connect the networks within this transnational movement and the deployment of already existing capacities, which are currently geared to counter Islamist terrorism online. (Topic of this briefing paper)
(3) More in-depth analytics concerning the financial activities and transnational commercial connections of the vXRW and terrorist movement to allow for the potential adjustment of existing global counter terrorism financing mechanisms.
(4) Greater awareness and the development of appropriate countermeasures focusing on the training activities within the vXRW movement, in particular paramilitary training.
(5) The further development of P/CVE approaches and concepts on a local, national and transnational level, based on lessons learned.

Throughout 2021, CEP, in cooperation with the Federal Foreign Office of Germany, will address these issues in a series of virtual events, bringing together relevant national and multilateral policy stakeholders. These events will be accompanied by a series of short reports, outlining the main operational and policy issues. Building on the discussions with relevant stakeholders, these papers will contain a range of concrete policy recommendations.

[1] Counter Extremism Project (CEP), Violent Right-Wing Extremism and Terrorism – Transnational Connectivity, Definitions, Incidents, Structures and Countermeasures, November 2020, (CEP 2020 study), https://www.counterextremism.com/sites/default/files/CEP%20Study_Violent%20Right-Wing%20Extremism%20and%20Terrorism_Nov%202020.pdf.

Misuse of Online Platforms by Violent Right-Wing Extremists and Terrorists

Challenges and potential countermeasures

The 2020 CEP study outlined that the transnational vXRW and terrorist movement has developed specific online ecosystems which are an integral part of the movement. In addition to meetings and gatherings in geographic network hubs, the online sphere plays a crucial role in the transnational functionality of the movement. This was/is particularly the case during the COVID-19 pandemic.[2]

Within this unorganized collective, individuals and groups feel connected by shared values, actions and perceived enemies. However, since this movement is leaderless and characterized by a variety of organizational models and structures as well as networks of loosely connected individuals,[3] the online ecosystem mirrors this variety and is characterized by a variety in online nodes. Different sub-milieus are, for instance, the transnational XRW Mixed Martial Arts (MMA) and music scenes, the accelerationists like Atomwaffen Division (AWD)/"lone" actors, the traditional groups/parties like Nordic Resistance Movement (NRM) or the Nationaldemokratische Partei Deutschlands (NPD). They have all developed their own distinct online communities, employing a wide range of strategies and tactics. Therefore, tailored analyses and appropriate countermeasures should be deployed.

In recent years, due to pressure from civil society and policymakers, Instagram and its parent company Facebook, along with YouTube and Twitter have started to remove some violence-oriented right-wing extremist actors and their content from their platforms. However, media reports indicate these efforts are less stringent than for extremist Islamist material.[4] Research shows that this "de-platforming" has led to a migration of these actors to smaller or alternative platforms like VKontakte, BitChute, meme/message-boards like 4Chan and 8Chan, and messenger services like WhatsApp or Telegram. The new accounts on those smaller platforms often have significantly fewer followers. This demonstrates the operational disruption and limitation of reach that can be achieved by removing violent XRW groups from global platforms.[5]

However, increased efforts in content moderation by social media platforms forced violence oriented right-wing extremist actors to modify their behaviour on global platforms, but it did not end their presence. Through an in-depth mapping exercise,[6] a new CEP research and analysis project produced an inventory of the most relevant actors of right-wing extremism in Germany. Based on criteria such as the number and relevance of activities (e.g., concerts, festivals, martial arts events, rallies, orientation towards violence, propaganda activities, national and transnational networking), a total of 100 individuals, organizations, music labels, bands,

[2] CEP 2020 study, p. 19.
[3] CEP 2020 study, p. 11f.
[4] CEP 2020 study, p. 29
[5] CEP 2020 study, p. 29.
[6] This CEP research report hasn't been published, yet. For an overview see: Kinetz, Erika, Neo-Nazis are still on Facebook. And they're making money, AP News, September 25, 2021, https://apnews.com/article/coronavirus-pandemic-lifestyle-technology-sports-business-43cc5e3bc9fbdd6d2d2f425c117e4f0a

fashion brands and companies were identified as (in many cases violence-oriented) key right-wing extremist actors.

A large proportion of these key actors in the various German right-wing extremist milieus are still present on Facebook, Instagram, YouTube and Twitter. However, generally, they no longer carry out illegal activities on these platforms. For the most part, they also do not violate the companies' general terms and conditions or community standards with regards to hate speech or similar offences in order to avoid being permanently blocked from the platforms.

These far-right extremists are now pursuing a strategy of "extreme normalization" on major social media. Nonetheless, they remain the same key right-wing extremist actors who form the foundation of Germany's transnational right-wing extremist scenes through festivals, fashion brands and mixed martial arts tournaments. Beyond their activities on major social media and video-sharing platforms, these key actors remain architects as well as advocates of dangerous conspiracy myths and narratives of group-based misanthropy that increase the likelihood of violence and terrorism.

Since many right-wing extremist key actors explicitly pursue economic interests, they use social media to promote their merchandise stores, martial arts associations, music labels, bands and survivalist (or prepper) organizations and to reach new customers and to recruit new followers and members.

Of the total 100 identified (violence-oriented) far-right key actors:

Facebook: 54 profiles can be assigned to 41 key actors, with a total of 267,743 subscribers/friends. Of these 54 profiles, 39 have explicit economic purposes.

Instagram: 37 profiles can be assigned to 34 key actors, with a total of 82,957 followers. Of these 37 profiles, 24 have explicit economic purposes.

YouTube: 33 profiles can be assigned to 27 actors, with a total of 82,160 subscribers. Of these 33 profiles, 15 have explicit economic purposes. The total number views for all videos is 9.594.861.

Twitter: 17 profiles of 16 actors were identified with a total of 6,818 followers. Of these 17 profiles, 9 have explicit economic purposes.

While a large part of known (violence-oriented) German far-right movement-entrepreneurs are present and active on Facebook, Instagram, YouTube and, to a much lesser extent, Twitter, big tech is to some extent engaged in countering hate speech, extremism and terrorism on their platforms. However, established policy dialogues that focus on the self-regulation of this industry as well as workshops with civil society groups are usually limited to harmful or illegal behaviour on the platforms themselves. As our research shows, however, key German (violence-oriented) right-wing extremist actors have adapted their strategies for propaganda, recruitment and funding to avoid being banned from popular social media platforms. Exposing this new strategy and suggesting solutions was the motivation for this report.

The question of whether social media platforms can be made to enforce their own community standards and delete profiles of, for example (violence-oriented) extremist actors or hate organizations, is discussed in detail in this paper. On the one hand, this would break new legal ground in Germany; on the other, some community standards promise a somewhat “safe space" to users which needs to be honoured.

Policy Options

To counter this malign misuse, a combination of effective internal industry defensive mechanisms guided by clear and effective regulatory frameworks are necessary.

Building on an analysis of these strategies and tactics, a range of preventative countermeasures by the industry could be employed:

A) Effective proactive moderation by platform providers based on increased awareness of the distinct roles that global platforms and smaller providers play as well as the structures of the various online ecosystems of the various violent right-wing extremist and terrorist networks.

B) Deployment of strategic communication also in messenger apps and gaming platforms, as well as tailored alternative or counter-narratives.

In addition, regulatory frameworks could be tailored more specifically towards this growing threat. In recent years a range of measures have been taken by stakeholders within the tech industry to counter the misuse of their services. The Christchurch Call for Action and the development of the Global Internet Forum to Counter Terrorism’s (GIFCT) Content Incident Protocol demonstrate this. However, significant failures, such as the failure to quickly contain the spread of the Christchurch attack video[7] or to effectively moderate online communities prior to the attack on the US Capitol in January 2021[8] continue to occur regularly.

Consequently, in addition to voluntary industry initiatives, government regulation is playing an indispensable role. Such regulation could:

A) create legal clarity concerning the obligations of platform providers

B) set clear and consistent standards for content moderation

C) require greater transparency of platforms in order to enable independent auditing of the effectiveness of the internal defence mechanisms within platforms.

D) provide commercial incentives through penalty systems which enable companies to strengthen their defensive mechanisms against abuse by extremist/terrorist actors.

In this regard, new legal developments such as the amendments of the German Network Enforcement Act (NetzDG) in 2020 and 2021, the passing of the Terrorism Content Online (TCO) Regulation of the European Union, as well as the ongoing negotiations concerning the Digital Services Act (DSA) of the European Union are crucial developments. In addition to necessary improvements of these new regulatory systems,[9] the effectiveness of these new

[7] Lapowsky, Issie, Why Tech Didn’t Stop the New Zealand Attack from Going Viral, Wired, 15 March 2019, https://www.wired.com/story/new-zealand-shooting-video-social-media/

[8] McEvoy, Jemima, Capitol Attack Was Planned Openly Online For Weeks—Police Still Weren’t Ready, Forbes, 7 January 2021, https://www.forbes.com/sites/jemimamcevoy/2021/01/07/capitol-attack-was-planned-openly-online-for-weeks-police-still-werent-ready/?sh=2cbe57ce76e2

[9] See for example:
Ritzmann, Alexander; Schindler, Hans-Jakob, NetzDG 2.0: Recommendations for the amendment of the German Network Enforcement Act (NetzDG), CEP, 2020,

regulatory frameworks will also depend on a common understanding between governments regarding the legal aspects of the threat and its connection to terrorism. For example, NetzDG and DSA require companies to remove "illegal" content after notification, while the TCO focuses on terrorism-related content.

https://www.counterextremism.com/sites/default/files/CEP%20NetzDG%202.0%20Policy%20Paper%20April%202020%20ENG.pdf;
Ritzmann, Alexander, Farid, Hany, Terrorist Content Online - How to build comprehensible transparency for automated decisionmaking systems (ADM), CEP, 2020, https://www.counterextremism.com/sites/default/files/CEP%20TCO%20ADM%20Transparency%202604.pdf;
Ritzmann, Alexander; Farid, Hany; Schindler, Hans-Jakob, The EU Digital Services Act (DSA): Recommendations for an Effective Regulation Against. Terrorist Content Online, CEP, 2020, https://www.counterextremism.com/sites/default/files/CEP%20Policy%20Paper_EU%20DSA_Sept%202020.pdf

ANNEX

Overview: European Union Regulatory Approach

EU Terrorism Content Online (TCO) Regulation

Overview[10]

The aim of the legislation is a swift removal of terrorist content online and to establish one common instrument for all member states to this effect. The rules will apply to hosting service providers offering services in the EU, whether or not they have their main establishment in the member states.

Voluntary cooperation with the hosting service providers will continue, but the legislation will provide additional tools for member states to enforce the rapid removal of terrorist content where necessary. Competent authorities in the member states will have the power to issue removal orders to the service providers, to remove terrorist content or disable access to it in all member states. The service providers will then have to remove or disable access to the content within one hour.

Hosting service providers exposed to terrorist content will need to take specific measures to address the misuse of their services and to protect their services against the dissemination of terrorist content. The decision as to the choice of measures remains with the hosting service provider.

The legislation also provides for a clear scope and a clear uniform definition of terrorist content in order to fully respect fundamental rights. It also includes effective remedies for both users whose content has been removed and for service providers to submit a complaint.

EU Digital Services Act (DSA)

CEP assessment of the draft legislation[11]

In September 2020, CEP contributed to the first round of consultations, particularly asking for the DSA to:

- Stop algorithmic amplification of illegal content
- Mandate clear and understandable transparency
- Learn from the NetzDG that the "notice and take down" mechanism is not effective
- Require proactive search for illegal content

[10] See: https://www.consilium.europa.eu/en/press/press-releases/2020/11/13/joint-statement-by-the-eu-home-affairs-ministers-on-the-recent-terrorist-attacks-in-europe/

[11] See: Alexander Ritzmann, Dr. Hans-Jakob Schindler and Lucinda Creighton, EU Commission consultation – Digital Services Act package – ex ante regulatory instrument of very large online platforms acting as gatekeepers, CEP, 2021, https://www.counterextremism.com/sites/default/files/2021-05/CEP%20Policy%20Paper%20II%20EU%20DSA_May%202021_0.pdf

- Learn from the regulatory and compliance structures (and failures) of the financial industry and others

The DSA draft published by the commission in December 2020 is an ambitious piece of legislation that aims, amongst other objectives, “to create a safer digital space in which the fundamental rights of all users of digital services are protected”.

Judging the draft legislation by this objective, CEP appreciates that stopping the algorithmic proliferation of illegal content and mandating understandable transparency (Art. 13, Art. 23, Art. 29 and Art. 33) are some of the core aspects the draft DSA aims at enforcing.

Unfortunately, the draft legislation does not address the continued failure of the existing “notice and action” moderation systems (Article 14) of gatekeepers. These have been highlighted repeatedly in studies and tests both by CEP and other organizations such as jugendschutz.net. On the contrary, the DSA perpetuates the “notice and action” mechanism as the main content moderation system, expecting the 400.000.000 internet users in the EU first to be exposed to illegal and possibly harmful content and then to notify the platforms about it. This means the externalization of safety and security functions to users rather than a requirement for platforms to ensure the safety of their customers or prevent harmful effects on the societies in which they conduct their commercial activities.

Article 19 pushes for gatekeepers to work with ‘trusted flaggers’ and other professional third parties, including governmental Internet Referral Units (IRUs), whose current mandate in most cases is limited to Islamist extremism/terrorism. While studies have shown that illegal content reported by trusted flaggers has a higher chance of being taken down or blocked than if the exact same illegal content is reported by users only*,* the DSA draft lacks clarity on how those trusted flaggers should be funded. At the moment, many trusted flagger organizations are civil society organizations that provide their expertise in the framework of small and short-term projects commissioned by governments. Those trusted flaggers mostly manually search for illegal content and are by no means capable of monitoring the billions of pieces of content that are produced by users each day on very large global platforms.

Lessons learned from CEP’s testing of gatekeepers “notice and take down” systems:

The draft also prohibits a general obligation (Article 7) to monitor information that intermediaries transmit or store and respectively does not mandate gatekeeper platforms to a proactive search for illegal content. As mentioned above, several evaluations and tests of the German NetzDG “notice and action” mechanisms demonstrate that neither users nor trusted flaggers are capable of providing the effective moderation needed to ensure that gatekeeper platforms are significantly safer. If the proposed content moderation approach contained in the current draft of the DSA remains unchanged, it is mostly left up to the decision of for-profit tech companies on how they may or may not protect their users.

During the CEP tests, two major lessons emerged:

Lesson 1: "notice and take down" systems might not work properly

For example, YouTube only blocked or deleted 35% of the reported manifestly illegal extremist/terrorist videos, despite the fact that the notice given by CEP included information about the official government ban orders which had been also confirmed by court decisions. Videos with identical content were blocked in some cases but not others, which further indicates that the applied system or process is defective or at minimum significantly inconsistent in its decision making.

- After a months-long process following the publication of CEP findings, Google Germany confirmed that they agreed with CEP's classifications of the reported content, which was then blocked by YouTube according to the NetzDG.

<u>Lesson 2: "notice and take down" systems, even if they work properly, might not achieve the objectives set by the current DSA proposals</u>

Facebook blocked manifestly illegal images reported by CEP but did not do so with unreported, manifestly illegal images in the same photo folder within the same Facebook account.

- Only after the publication of CEPs findings, Facebook removed the account in total. This indicates that "notice and takedown" is being implemented only in the narrowest possible manner and does not present an early warning system for the company, leading to more strategic defences against remaining manifestly illegal content, even when present within the same user account.

The DSA proposes new transparency requirements and external audits, which CEP very much welcomes in general. The lack of procedural clarity, however, might make those insufficient. Particularly, lessons from (failed) auditing/compliance systems for example from the financial sector, are largely ignored but should be at the centre of a DSA auditing framework.

Unfortunately, the DSA is still based on, and helps to continue spreading, a set of misinformed narratives about the status quo of the role, function and business models of so-called gatekeeper platforms. This paper will therefore first try to address those systemic misunderstandings and provide an alternative narrative that might help to build the internet and intermediary services the EU is actually aiming for. Secondly, we are sharing concrete learnings from the financial industry that should be implemented within the existing DSA draft.

Overview: Self-Regulatory Approaches

EU Internet Forum[12]

The EU Internet Forum is one of the key commitments made in the European Agenda on Security presented by the Commission in April 2015. It aims to provide a framework for an efficient cooperation with the internet industry in the future and to secure a commitment from the main actors to coordinate and scale up efforts in this area in the coming years. Senior representatives of Ask.fm, Facebook, Google, Microsoft and Twitter participated in the launch event of the EU Internet Forum.

The European Agenda on Security prioritises terrorism, organised crime and cybercrime as interlinked areas with a strong cross-border dimension, where EU action can make a real difference. One of these priorities is the development of effective de-radicalisation and disengagement programmes, including the influence of online radicalisation.

To enhance and support efforts to prevent radicalisation leading to violent extremism the Commission has established the Centre of Excellence at the Radicalisation Awareness Network (RAN Centre) to reinforce its anti-radicalisation work. The RAN Centre facilitates the exchange of best practices and expertise, consolidates knowledge and identifies and develops best practices, concrete guidance and tailor made support services.

The Commission hosted the first high-level conference on the criminal justice response to radicalisation on 16 October 2015. All Justice Ministers present insisted that one of the most worrying sources of radicalisation is currently online. They called for a revision of its detection and prevention methods.

EU Internet Forum: EU-wide Crisis Protocol 2019[13]

In the aftermath of the terror attack in Christchurch, New Zealand, government leaders and online platforms agreed on the Christchurch Call for Action. On this occasion, the then EU Commission President Juncker announced the development of an EU Crisis Protocol in the context of the EU Internet Forum. The EU Protocol will allow Member States and online platforms to respond rapidly and in a coordinated manner to the dissemination of terrorist content online in the event of a terrorist attack.

The EU Crisis Protocol endorsed by the EU Internet Forum will:

- Provide a coordinated and rapid reaction: Member States' authorities, together with Europol, the Global Internet Forum to Counter Terrorism (GIFCT) and online service providers will be able to respond quickly, in a coordinated manner to ensure that the spread of terrorist or violent extremist content is swiftly contained.
- Facilitate public and private sector cooperation: In the event of a crisis, law enforcement authorities and online service providers will share relevant information on the online content (e.g., URLs, audio-visual media, and metadata) on a voluntary basis, in a secure way and in real time.

[12] See: https://ec.europa.eu/commission/presscorner/detail/en/IP_15_6243

[13] See: https://ec.europa.eu/commission/presscorner/detail/en/IP_19_6009

- Facilitate a voluntary arrangement: The Protocol does not replace national legal frameworks or existing national crisis management mechanisms. It should apply only to extraordinary situations where those national measures are no longer sufficient to coordinate a rapid and cross-border response.

European Commission and IT Companies: Code of Conduct on illegal online hate speech[14]

The IT Companies (Facebook, Twitter, YouTube and Microsoft[15]), taking the lead on countering the spread of illegal hate speech online, agreed with the European Commission on a code of conduct setting the following public commitments:

- The IT Companies to have in place clear and effective processes to review notifications regarding illegal hate speech on their services so they can remove or disable access to such content. The IT companies to have in place Rules or Community Guidelines clarifying that they prohibit the promotion of incitement to violence and hateful conduct.
- Upon receipt of a valid removal notification, the IT Companies to review such requests against their rules and community guidelines and where necessary national laws transposing the Framework Decision 2008/913/JHA, with dedicated teams reviewing requests.
- The IT Companies to review the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary.
- In addition to the above, the IT Companies to educate and raise awareness with their users about the types of content not permitted under their rules and community guidelines. The use of the notification system could be used as a tool to do this.
- The IT companies to provide information on the procedures for submitting notices, with a view to improving the speed and effectiveness of communication between the Member State authorities and the IT Companies, in particular on notifications and on disabling access to or removal of illegal hate speech online. The information is to be channelled through the national contact points designated by the IT companies and the Member States respectively. This would also enable Member States, and in particular their law enforcement agencies, to further familiarise themselves with the methods to recognise and notify the companies of illegal hate speech online.
- The IT Companies to encourage the provision of notices and flagging of content that promotes incitement to violence and hateful conduct at scale by experts, particularly via partnerships with CSOs, by providing clear information on individual company Rules and Community Guidelines and rules on the reporting and notification processes. The IT Companies to endeavour to strengthen partnerships with CSOs by widening the

[14] See: https://ec.europa.eu/commission/presscorner/detail/en/IP_16_1937

[15] Since the establishment of the Code of Conduct in 2016, additional platform providers have joined. In the course of 2018, Instagram, Snapchat and Dailymotion took part to the Code of Conduct, Jeuxvideo.com in January 2019, and TikTok joined in September 2020. On 25 June 2021, LinkedIn also announced its participation to the Code of Conduct, see: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#theeucodeofconduct

geographical spread of such partnerships and, where appropriate, to provide support and training to enable CSO partners to fulfil the role of a "trusted reporter" or equivalent, with due respect to the need of maintaining their independence and credibility.

- The IT Companies rely on support from Member States and the European Commission to ensure access to a representative network of CSO partners and "trusted reporters" in all Member States helping to help provide high quality notices. IT Companies to make information about "trusted reporters" available on their websites.
- The IT Companies to provide regular training to their staff on current societal developments and to exchange views on the potential for further improvement.
- The IT Companies to intensify cooperation between themselves and other platforms and social media companies to enhance best practice sharing.
- The IT Companies and the European Commission, recognising the value of independent counter speech against hateful rhetoric and prejudice, aim to continue their work in identifying and promoting independent counter-narratives, new ideas and initiatives and supporting educational programs that encourage critical thinking.
- The IT Companies to intensify their work with CSOs to deliver best practice training on countering hateful rhetoric and prejudice and increase the scale of their proactive outreach to CSOs to help them deliver effective counter speech campaigns. The European Commission, in cooperation with Member States, to contribute to this endeavour by taking steps to map CSOs' specific needs and demands in this respect.
- The European Commission in coordination with Member States to promote the adherence to the commitments set out in this code of conduct also to other relevant platforms and social media companies.

Christchurch Call for Action[16]

Commitments made by participating Governments

- Counter the drivers of terrorism and violent extremism by strengthening the resilience and inclusiveness of our societies to enable them to resist terrorist and violent extremist ideologies, including through education, building media literacy to help counter distorted terrorist and violent extremist narratives, and the fight against inequality.
- Ensure effective enforcement of applicable laws that prohibit the production or dissemination of terrorist and violent extremist content, in a manner consistent with the rule of law and international human rights law, including freedom of expression.
- Encourage media outlets to apply ethical standards when depicting terrorist events online, to avoid amplifying terrorist and violent extremist content.
- Support frameworks, such as industry standards, to ensure that reporting on terrorist attacks does not amplify terrorist and violent extremist content, without prejudice to responsible coverage of terrorism and violent extremism.

[16] See: https://www.christchurchcall.com/call.html

- Consider appropriate action to prevent the use of online services to disseminate terrorist and violent extremist content, including through collaborative actions, such as:
  - Awareness-raising and capacity-building activities aimed at smaller online service providers;
  - Development of industry standards or voluntary frameworks;
  - Regulatory or policy measures consistent with a free, open and secure internet and international human rights law.

Commitments made by participating online service providers

- Take 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, without prejudice to law enforcement and user appeals requirements, in a manner consistent with human rights and fundamental freedoms. Cooperative measures to achieve these outcomes may include technology development, the expansion and use of shared databases of hashes and URLs, and effective notice and takedown procedures.
- Provide greater transparency in the setting of community standards or terms of service, including by:
  - Outlining and publishing the consequences of sharing terrorist and violent extremist content;
  - Describing policies and putting in place procedures for detecting and removing terrorist and violent extremist content.
- Enforce those community standards or terms of service in a manner consistent with human rights and fundamental freedoms, including by:
  - Prioritising moderation of terrorist and violent extremist content, however identified;
  - Closing accounts where appropriate;
  - Providing an efficient complaints and appeals process for those wishing to contest the removal of their content or a decision to decline the upload of their content.
- Implement immediate, effective measures to mitigate the specific risk that terrorist and violent extremist content is disseminated through livestreaming, including identification of content for real-time review.
- 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.
- Review the operation of algorithms and other processes that may drive users towards and/or amplify terrorist and violent extremist content to better understand possible intervention points and to implement changes where this occurs. This may include using algorithms and other processes to redirect users from such content or the promotion of credible, positive alternatives or counter-narratives. This may include building appropriate mechanisms for reporting, designed in a multi-stakeholder process and without

compromising trade secrets or the effectiveness of service providers’ practices through unnecessary disclosure.

- Work together to ensure cross-industry efforts are coordinated and robust, for instance by investing in and expanding the GIFCT, and by sharing knowledge and expertise.

Tech Against Terrorism[17]

Tech Against Terrorism is an initiative launched and supported by the United Nations Counter Terrorism Executive Directorate (UN CTED) working with the global tech industry to tackle terrorist use of the internet whilst respecting human rights.

Its plan of action revolves around three pillars: outreach, knowledge-sharing, and practical support. Tech Against Terrorism promotes constructive working relationships between the tech and government sectors and organises global workshops and e-learning sessions to conduct in-person training with tech companies. Tech Against Terrorism works with the global tech sector to share best practice (policy, guidelines, learning materials, practical workshops, and tools) within the tech industry and across the private, public, and civil society sectors.

[17] See: https://www.techagainstterrorism.org/about/

Overview: industry led-mechanisms

Global Internet Forum to Counter Terrorism (GIFCT)[18]

Mission and Vision

The Global Internet Forum to Counter Terrorism (GIFCT) is an NGO designed to prevent terrorists and violent extremists from exploiting digital platforms. Founded by Facebook, Microsoft, Twitter, and YouTube in 2017, the Forum was established to foster technical collaboration among member companies, advance relevant research, and share knowledge with smaller platforms. Since 2017, GIFCT's membership has expanded beyond the founding companies to include over a dozen diverse platforms committed to cross-industry efforts to counter the spread of terrorist and violent extremist content online.

These efforts have evolved in conjunction with the Christchurch Call to Action, an initiative that governments, tech platforms, and civil society organizations committed to after the March 2019 Mosque shootings in Christchurch, New Zealand and viral spread of the perpetrator's live-streamed video of the attack. In addition to the Christchurch Call, tech companies also signed onto a nine-point plan designed to support industry efforts to eliminate terrorist and violent extremist content online. At an UNGA side event led by New Zealand Prime Minister Jacinda Ardern and French President Emmanuel Macron in September 2019, the founding companies announced that GIFCT would spin off as an independent 501(c)(3) with its own dedicated technology, counterterrorism, and operations teams.

Four foundational goals guide the newly independent organization's work:

1. Empower a broad range of technology companies, independently and collectively, with processes and tools to prevent and respond to abuse of their platforms by terrorists and violent extremists
2. 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
3. Promote civil dialogue online and empower efforts to direct positive alternatives to the messages of terrorists and violent extremists
4. Advance broad understanding of terrorist and violent extremist operations and their evolution, including the intersection of online and offline activities.

[18] See: https://gifct.org/about/