Preventing Terrorist Use of the Internet

The Role for Commonwealth Governments

The Commonwealth

Acknowledgements This publication was prepared by Tech Against Terrorism

Jacob Berntsson
Arthur Bradley
Anne Craanen
Adam Hadley
Maygane Janin
Deeba Shadnia
Fabienne Tarrant

For further information on this report contact: Countering Violent Extremism Unit, Governance and Peace Directorate, Commonwealth Secretariat Marlborough House Pall Mall London SW1Y 5HX United Kingdom Tel: +44 (0)20 7747 6534 Email: cve@commonwealth.int Web: https://thecommonwealth.org/countering-violent-extremism

.

WORKSHOP RESOURCES FROM THE

Commonwealth Technical Assistance Workshop: Preventing Terrorist Use of the Internet – The Role for Commonwealth Governments

Africa Regional Workshop

27 September – 1 October 2012

The Commonwealth

Contents

Welcome Message 1
- Mark Albon Head of Countering Violent Extremism Unit, Commonwealth Secretariat 1

Dossier A: Global and Sub-Saharan Africa: Terrorist use of the internet 2
- Executive Summary 2
- Internet Use in Sub-Saharan Africa 2
- Terrorist Use of the Internet: Overview 6
- Terrorist Use of the Internet: Strategic and Operational Purposes 12
- What makes a platform attractive to terrorists? 16
- Key trends in terrorist use of the Internet 2
- Differences in the use of the internet across terrorist and violent extremist ideologies 6
- Methodology for OSINT Investigations 7
- Case Study: Ansar al-Sunna 8
- Case Study: Boko Haram 21
- Case Study: Al-Shabaab 28

Dossier B: Legal Responses to Terrorist Use of the Internet 38
- Executive Summary 38
- Global trends in online regulation and legal responses to terrorist use of the internet: 39
- Tech Against Terrorism’s Recommendations for Governments 40
- Emerging Online Regulation – Good Practice 41
- Emerging Online Regulation – Key Trends and Concerns 44
- Legal Responses to Terrorist Use of the Internet: Africa 46

Dossier C: Public-Private Partnerships and Cross-Industry Cooperation 51
- Executive Summary 51
- Tech Against Terrorism: Public-Private Partnership 52
- Existing public-private partnerships aimed at tackling terrorist use of the internet 58
- Gap analysis: public-private partnerships and terrorist use of the internet 63

Dossier D: Tech Sector Efforts to Counter Terrorism and Cross-Industry Cooperation 65
- Executive Summary 65
- Existing industry cooperation mechanisms 66
- Tech Sector Efforts 71

Dossier E: Countering terrorist use of the internet whilst respecting human rights 90
- Executive Summary 90
- Online Counter-Terrorism: Summary of Human Rights Risks 91
- Online Counter-Terrorism & Human Rights: Regional Case Studies 101

Dossier F: Removing Terrorist Content 109
- Executive Summary 109
- The Terrorist Content Analytics Platform (TCAP) 110
- Terrorist content: definitions and identification challenges 122

Welcome Message

Mark Albon
Head of Countering Violent Extremism Unit, Commonwealth Secretariat

“It is imperative to counter the use of the internet by extremist groups to radicalise and recruit”

The Commonwealth Cyber Declaration recognises that, since its inception, the internet has made a powerful contribution to the economic, social, cultural and political life of the Commonwealth. But extremist groups’ use of the internet is a threat to global security, peace and stability. In 2015, Commonwealth Heads of Government agreed that it is imperative to counter the use of the internet by extremist groups to radicalise and recruit.

The Commonwealth Secretariat is pleased to be working across the Commonwealth to support its members in preventing terrorist misuse of the internet and social media platforms. The provision of advisory services to Commonwealth Governments on countering violent extremism is a core activity of the Governance and Peace Directorate.

We maintain in-house expertise to provide support on countering violent extremism and terrorism prevention and are pleased to hear from external experts, including the Commonwealth Telecommunications Organisation and Tech Against Terrorism, to help Commonwealth members understand and respond to the threat of terrorist use of the internet.

Responding to terrorist use of the internet requires us to balance the competing calls for security, enforcement, freedom of expression, transparency and accountability. When this issue is viewed through only one of these lenses, we can only ever go part-way to solving the problem. What is needed is a robust approach that holds all of these competing needs and navigates a way forward.

The Commonwealth Secretariat is pleased to convene this virtual workshop so that policymakers, regulators, law enforcement officers, the tech sector and civil society representatives can meet with global experts from tech sector, research and civil society to discuss the ways that terrorists use the internet (regionally and globally), and to explore positive and sensible measures that Commonwealth members can take to reduce the proliferation of violent extremist materials online, while upholding the values articulated in the Commonwealth Charter.

I wish you all fruitful discussions at the Workshops.

Dossier A: Global and Sub-Saharan Africa: Terrorist use of the internet

Executive Summary

This research dossier presents Tech Against Terrorism's research on terrorist use of the internet across Sub-Saharan Africa, using three case studies on different regional terrorist actors: Ansar al-Sunna, Boko Haram, and al-Shabaab. While these three terrorist groups all adhere to violent Islamist ideologies, their strategic and operational objectives differ, and there are distinctions in the nature and extent of how they operate online.

These case studies highlight that terrorist actors in Sub-Saharan Africa exploit the internet in a variety of ways depending on several factors, including affiliation with other terrorist groups, operational capacity, and prevalence of local internet access.

Understanding terrorist use of the internet in Sub-Saharan Africa must be contextualised within general trends of internet use and access in the region. Internet access across the African continent varies between regions, however overall, there have been significant increases to public internet access in recent years.

Despite this improvement, only 18% of the continent's population has regular access to the internet, compared with a global average of 30%. On the other hand, the wide use of mobile phone technologies across Sub-Saharan Africa has dramatically impacted communications, as well as the use of social media across the region.

This dossier provides background information on how terrorist and violent extremist (T/VE) actors use the internet, as well as current trends in terrorist use of the internet. In general, T/VE actors exploit the internet for strategic and operational purposes including recruitment, fundraising and the dissemination of propaganda.

Through continued research and monitoring, Tech Against Terrorism assesses that T/VE actors exploit a multitude of different sites and platforms to ensure their content remains online for as long as possible. This strategy particularly raises challenges for small tech platforms, who often do not have the capacity or resources to fully counter the threat.

This report notes that among several recent trends in terrorist use of the internet, the most significant are the recent resurgence of terrorist operated websites, and the ongoing development of operational security awareness among terrorist actors.

Terrorist operated websites (TOWs) present several challenges to law enforcement and counter-terrorism practitioners, as the removal of these sites often takes a long time. Additionally, it is not always clear who is legally responsible for the website's removal.

In recent years T/VE actors across the ideological spectrum have developed a sophisticated awareness of how to evade detection by law enforcement and content moderators through several ways. These include the use of VPNs, Tor browsers and methods at evading automated moderation tools on social media platforms.

Internet Use in Sub-Saharan Africa

To adequately assess how terrorist and violent extremist (TVE) actors in Sub Saharan Africa (SSA) are exploiting online spaces, we must first understand how people in the region use the internet.

Access to the internet across SSA has improved in recent years, although public access to the internet in Sub-Saharan Africa remains lower than global access.[1]

[1] https://www.cfr.org/blog/last-month-over-half-billion-africans-accessed-internet

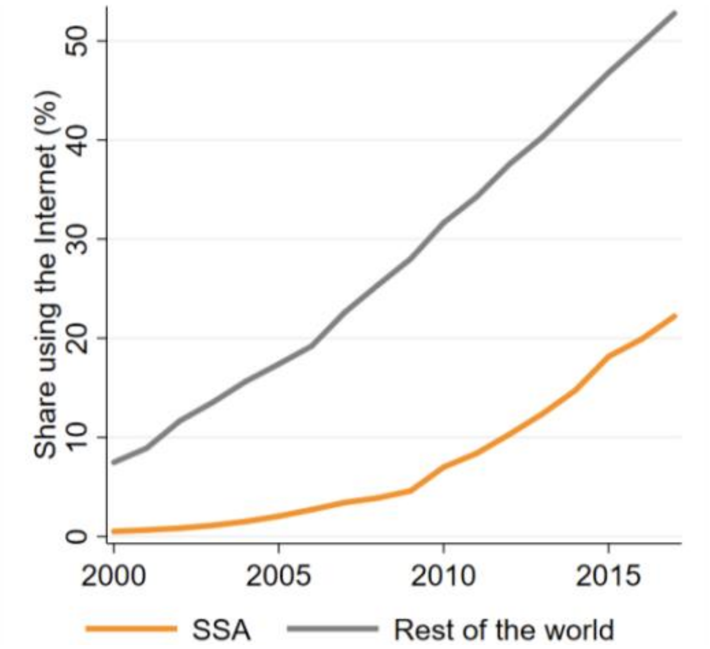


Figure 1: Internet usage in Sub Saharan Africa and the rest of the world. Source: The World Bank.[2]

There are estimated to be approximately 170 million internet users across the continent, which indicates that 18% of Africa's population has access to the internet. This is significantly lower than the global average of 30%. Approximately one in ten households is connected to the internet.[3] According to the International Telecommunications Union (ITU), which tracks internet usage globally and across countries, only 1 in 5 members of the public in SSA used the internet in 2017.[4]

Within SSA, the rate of internet usage varies between countries and regions. In Southern Africa, nearly half the population uses the internet, while in West Africa the proportion is closer to 30%. In Central Africa, by contrast, only 10% of the population is recorded as having access to the internet.[5] The Pew Research Centre reported in January 2018 that 25% of respondents in Tanzania had access to the internet, compared with 89% in the US. Research has noted that rates of internet usage are particularly low in landlocked countries across SSA, where there is often a shortage of the physical infrastructure needed to facilitate internet access. [6]

[2] https://documents1.worldbank.org/curated/en/518261552658319590/pdf/Internet-Access-in-Sub-Saharan-Africa.pdf

[33] https://interactive.aljazeera.com/aje/2016/connecting-africa-mobile-internet-solar/internet-connecting-africa.html

[4] https://documents1.worldbank.org/curated/en/518261552658319590/pdf/Internet-Access-in-Sub-Saharan-Africa.pdf

[5] https://documents1.worldbank.org/curated/en/518261552658319590/pdf/Internet-Access-in-Sub-Saharan-Africa.pdf

[6] https://documents1.worldbank.org/curated/en/518261552658319590/pdf/Internet-Access-in-Sub-Saharan-Africa.pdf

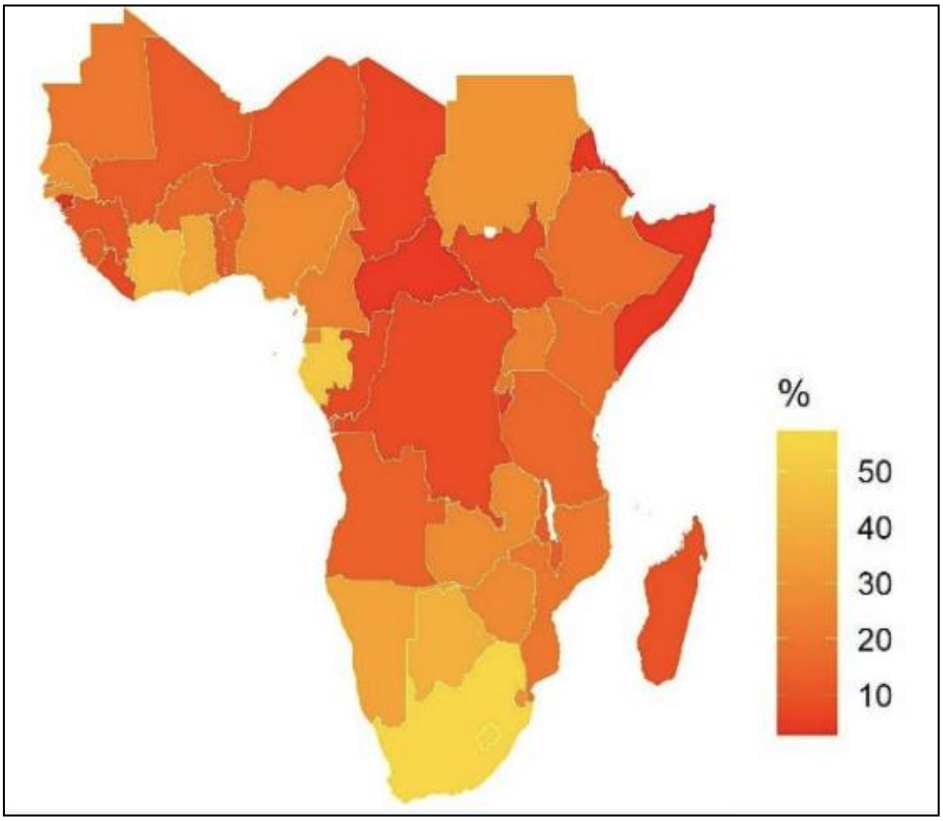


Figure 2: Internet use as a percentage of country population in Sub Saharan Africa, 2017.[7]

Access to broadband internet across SSA is expanding at a relatively slow rate compared to other regions, but mobile usage is more widespread than electricity.[8] In 2016, The Economist reported that while less than half the population in SSA had access to electricity, two fifths owned a mobile phone.[9] The expansion of mobile phone technologies has been crucial to the region, as mobiles are often the sole or primary method of digital access for many individuals. In 2017, for every 100

[7] https://documents1.worldbank.org/curated/en/518261552658319590/pdf/Internet-Access-in-Sub-Saharan-Africa.pdf

[8] https://www.economist.com/graphic-detail/2017/11/08/in-much-of-sub-saharan-africa-mobile-phones-are-more-common-than-access-to-electricity

[9] https://www.economist.com/graphic-detail/2017/11/08/in-much-of-sub-saharan-africa-mobile-phones-are-more-common-than-access-to-electricity

people in people in the region, there were 34 active mobile broadband subscriptions, compared to just 0.4 for fixed broadband.[10]

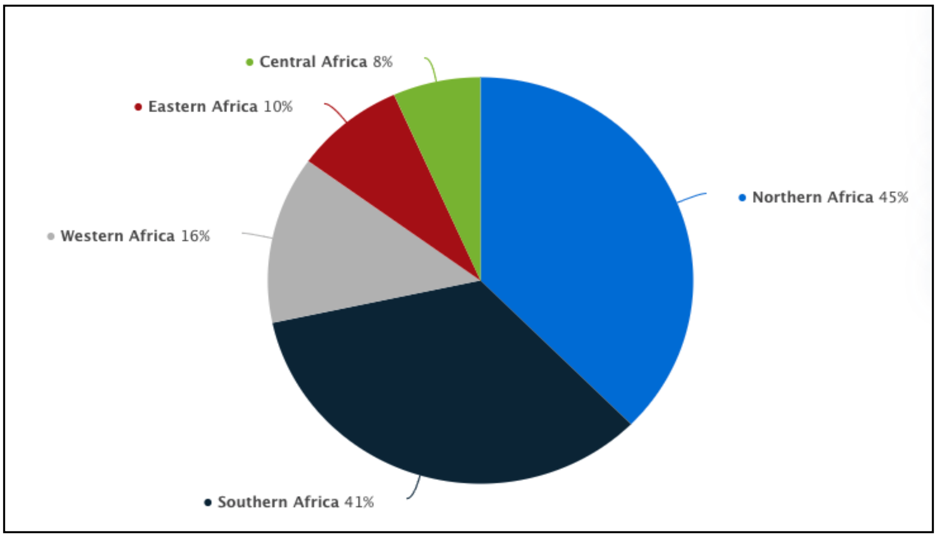


Figure 3: Social media penetration in Africa in 2021, by region.[11]

As the use of mobile networks has grown significantly across SSA, so too has the use of social media. Social media use is especially prevalent among youth, and SSA has a high youth population. In 2018, Pew Research Centre noted that more than three quarters of Sub-Saharan Africans who go online also use social media.[12] The same report said: "Social media users are much more likely to use these sites to share their views about entertainment topics than about other issues. For example, majorities of social media users say they use social media to share their views about music and movies (61%) and sports (57%). Far fewer post about religion (45%), politics (37%) or products they use (37%)".[13]

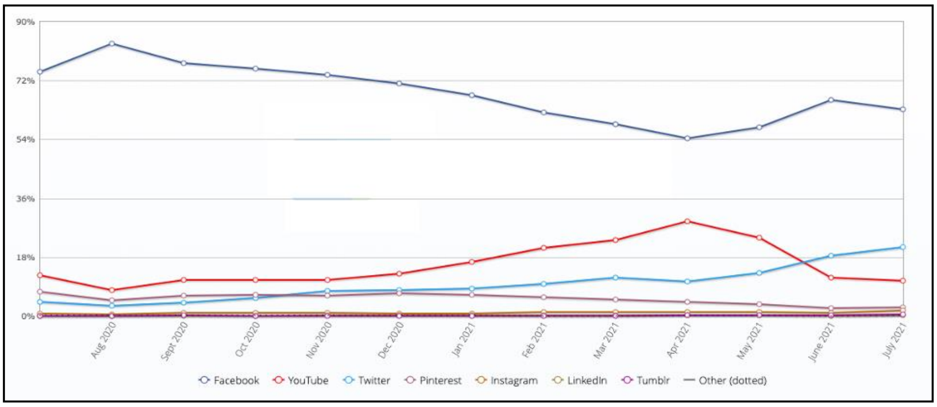


Figure 4: Percentage of population using social media platforms across Africa in July 2020 – July 2021.

[10] https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2019.pdf

[11] https://www.statista.com/statistics/1190628/social-media-penetration-in-africa-by-region/

[12] https://www.pewresearch.org/global/2018/10/09/internet-use-is-growing-across-much-of-sub-saharan-africa-but-most-are-still-offline/

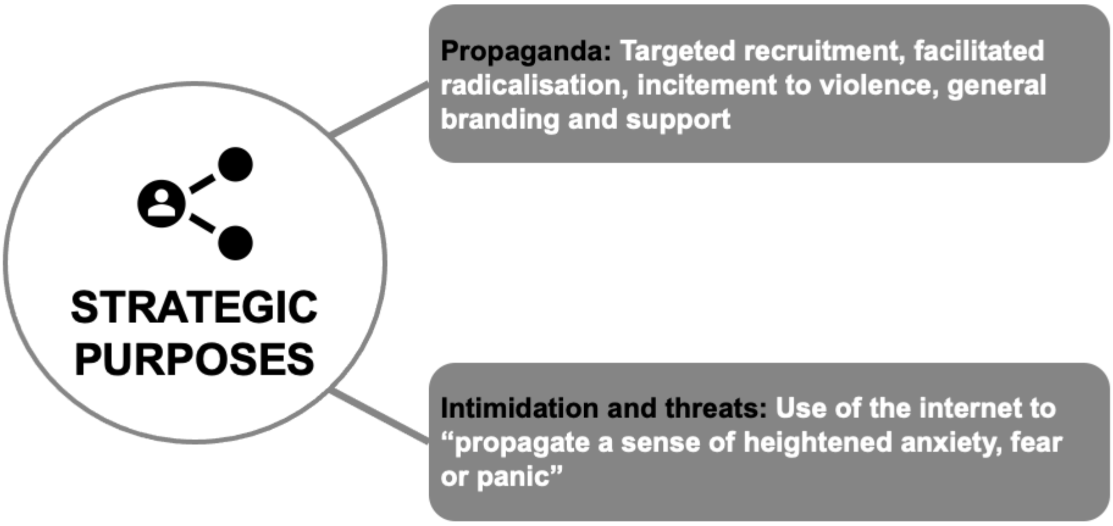
[13] https://www.pewresearch.org/global/2018/10/09/internet-use-is-growing-across-much-of-sub-saharan-africa-but-most-are-still-offline/

Terrorist Use of the Internet: Overview

Terrorists and violent extremists (T/VE) have long made use of the internet to spread propaganda and to communicate internally. Terrorist groups from all forms of violent extremism, including racially- and ethnically motivated violent extremism groups (such as far-right VEOs) and religious-motivated violent extremism groups (such as Islamist VEOs) occupy complex and wide-reaching online ecosystems, and their presence now spans a broad range of platforms.

Tech Against Terrorism assesses that in general, T/VE actors exploit the internet for a variety of different strategic purposes and operational purposes.

The strategic purposes of the internet for T/VE actors are served by the ways in which the internet sustains the durability and growth of T/VE entities by facilitating, for example, the dissemination of narratives supportive of a cause, the sanitisation of their public image via propaganda, and the attempted recruitment of others.

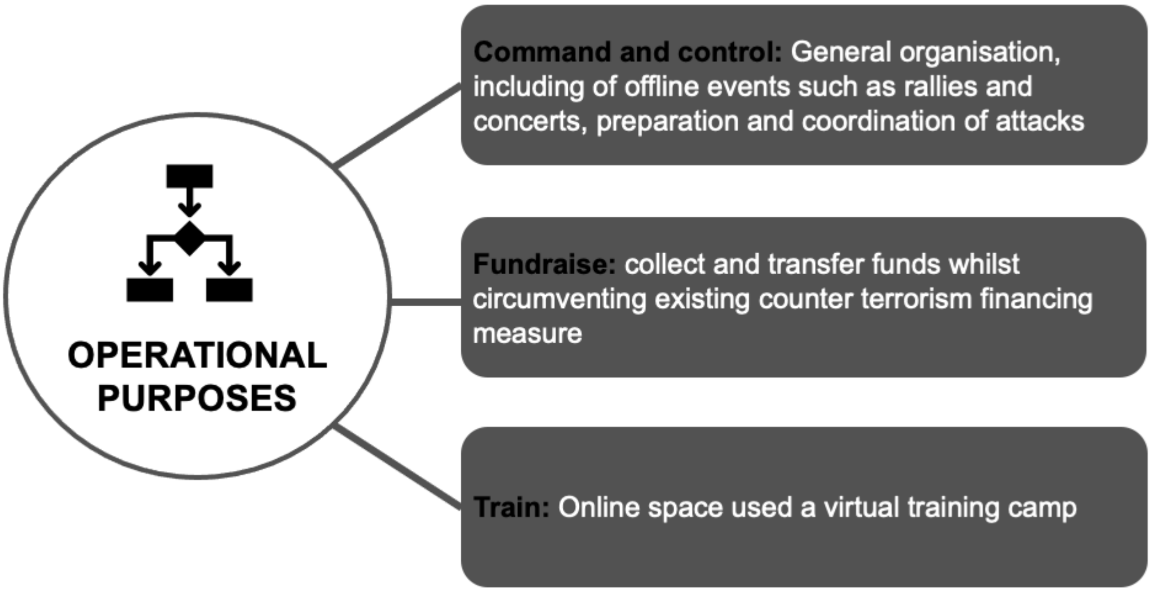


STRATEGIC PURPOSES

Propaganda: Targeted recruitment, facilitated radicalisation, incitement to violence, general branding and support

Intimidation and threats: Use of the internet to "propagate a sense of heightened anxiety, fear or panic"

The operational purposes of T/VE use of the internet are served by facilitated day-to-day management and operational organisation of a T/VE group. Examples include internal communication efforts, training and fundraising.



OPERATIONAL PURPOSES

Command and control: General organisation, including of offline events such as rallies and concerts, preparation and coordination of attacks

Fundraise: collect and transfer funds whilst circumventing existing counter terrorism financing measure

Train: Online space used a virtual training camp

T/VE actors attempt in particular to maintain a presence on mainstream social media platforms and do so despite the platforms' moderation efforts. The attractiveness of these platforms lies in their large userbases, which terrorists and violent extremists seek in order to broadcast their content to as wide an audience as possible, and to aid recruitment efforts.

Multiplatform dissemination strategies

T/VE actors rely on a multiplatform approach to ensure both the rapid sharing of content and a resilient online presence.[14] This mitigates the impact of moderation and takedown attempts by diversifying the bases and thereby building the resilience of their online presence, as well as by maximising the rate at which their content can be disseminated.

[14] Fisher Ali, Prucha Nico, Winterbotham Emily, (2019), "Mapping the Jihadist information ecosystem: Towards the next generation of disruption capability", *Global Research Network on Terrorism and Technology,* Paper No. 6

Types of platforms used by terrorists and violent extremists

Platform Type	Offering	Examples
Social Media Platforms	Social media platforms offer terrorists the opportunity to reach the widest possible audience, and to have engage in dialogue with their members and supporters.	Facebook, Twitter, VKontakte, Instagram
Messaging Applications	Messaging apps offer T/VE an easy navigable, free, and often secure means of both internal and external communication. Most messaging apps are protected by either end-to-end or client-server encryption.	WhatsApp, Telegram, Hoop Messenger, WeChat, Line, TamTam, Viber, Slack
Alt-tech platforms	A variety of platforms have emerged in the past few years that claim to offer an alternative to larger, more mainstream platforms like Twitter, YouTube and Facebook. These platforms often explicitly market themselves as ‘free speech’ platforms, or ones that oppose the ‘internet censorship’ of larger platforms’ content moderation policies. Some alt-tech platforms use decentralised blockchain-based technology, which makes content moderation more difficult by removing centralised administrators.	Gab, Parler, Bitchute, Minds, RocketChat, Mastodon
Video Sharing Platforms (VSPs)	VSPs provide terrorists and violent extremists with an ideal platform on which to promote their audio-visual content. Search functions within these sites mean that content can easily be found, and file size limits are typically larger than on most other online platforms.	YouTube, Bitchute, DLive, Vimeo, DailyMotion LiveLeak, Veoh, TikTok
Pasting Sites	Pasting sites are used by terrorists and violent extremists to store content such as videos, images and audio files. They are also used to aggregate information, such as lists of URLs to further content.	Justpaste.it, DropBox, Archive.org, Top4Top, Zippyshare, Files.fm,
Gaming Platforms	Terrorists and violent extremists use gaming platforms to propagate their ideologies and to recruit through video games. They have also used gaming platforms to communicate and plan attacks and other forms of political violence, as well as to stream attacks.	Twitch, Discord, Steam, Roblox
Audio Streaming Platforms	Terrorists and violent extremists exploit audio streaming platforms to share voice messages, extremist music, and audiobook versions of written documents such as terrorist manifestos.	SoundCloud, Spotify, Apple music, BandCamp
Terrorist Operated Websites	Websites that are run by terrorist groups and their supporters dedicated to a terrorist group’s interests. These play an important role in the online terrorist ecosystem, often acting as a centralised hub of content that may have been removed from social media platforms and messaging channels.	Sahab, Shahada News Agency, Obedient Supporters, Elokab, Emaad, Al-Bayaan

Use of online platforms by terrorists and violent extremists

A 2019 study by Ali Fisher, Nico Prucha, and Emily Winterbotham identifies three main uses of the above types of platforms by terrorists online. These are categorised in terms of beacons, content stores and aggregators.[15] To this, Tech Against Terrorism has added a fourth category: circumventors. These are summarised below.

[15] Fisher Ali, Prucha Nico, Winterbotham Emily, (2019), 'Mapping the Jihadist Information Ecosystem: Towards the Next Generation of Disruption Capability', *Global Research Network on Terrorism and Technology*, Paper No. 6, available at: https://rusi.org/sites/default/files/20190716_grntt_paper_06.pdf

BEACONS

Platforms used by terrorists and violent extremists to project their content to the widest audience possible. The beacon acts both as a centrally located lighthouse and signpost to where the content can be found. Through beacons, terrorists redirect their target audience to the platforms on which content is hosted.

CONTENT STORES

Where terrorist content is stored, including text and audio files, as well as images and videos. These are used as online libraries of content. Terrorists and violent extremists rely on content storage platforms and pasting sites, as well as archive services.

AGGREGATORS

Aggregators act as centralised databases of where content can be found online, gathering together a wide range of URLs to content hosting platforms to facilitate diffusion. If one link is taken down, terrorists can easily find an alternative to share.

CIRCUMVENTORS

Online services and platforms used to circumvent content moderation and deplatforming measures. Circumventors include VPNs, which can enable nefarious actors to access content that has been blocked in specific countries. Another example of circumventors is the use of decentralised web technologies, which avoid website takedowns. In our analysis, Discord fits into both the Beacon and Aggregator category; based on Tech Against Terrorism's research so far terrorists and violent extremists have used Discord servers both as a centrally located signpost for content and communication, as well as a place where they can upload and store content.

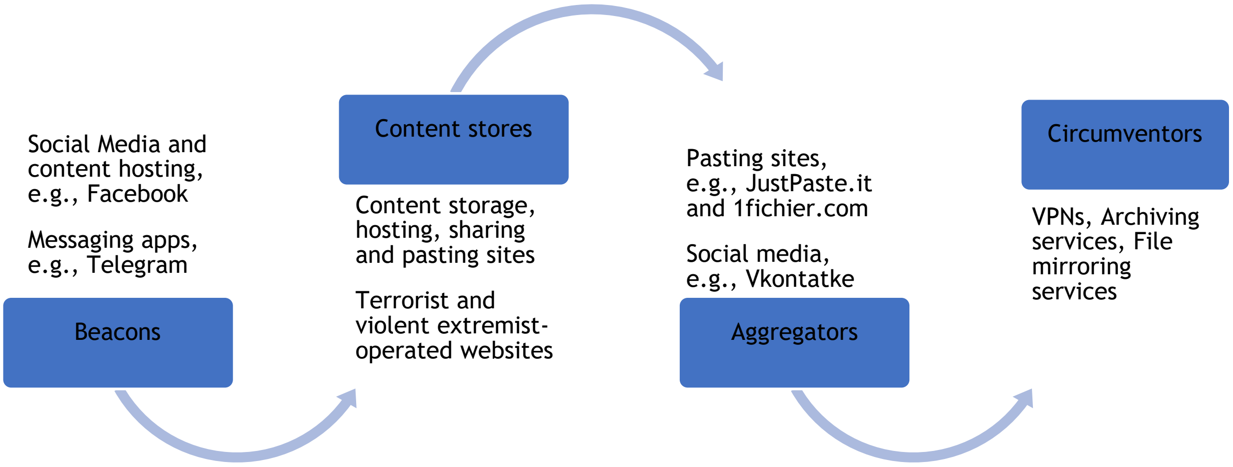

Social Media and content hosting, e.g., Facebook
Messaging apps, e.g., Telegram
Beacons
Content stores
Content storage, hosting, sharing and pasting sites
Terrorist and violent extremist-operated websites
Pasting sites, e.g., JustPaste.it and 1fichier.com
Social media, e.g., Vkontatke
Aggregators
Circumventors
VPNs, Archiving services, File mirroring services

Figure 5: Eco-system of online platforms most commonly used by terrorists and violent extremists in a multiplatform approach

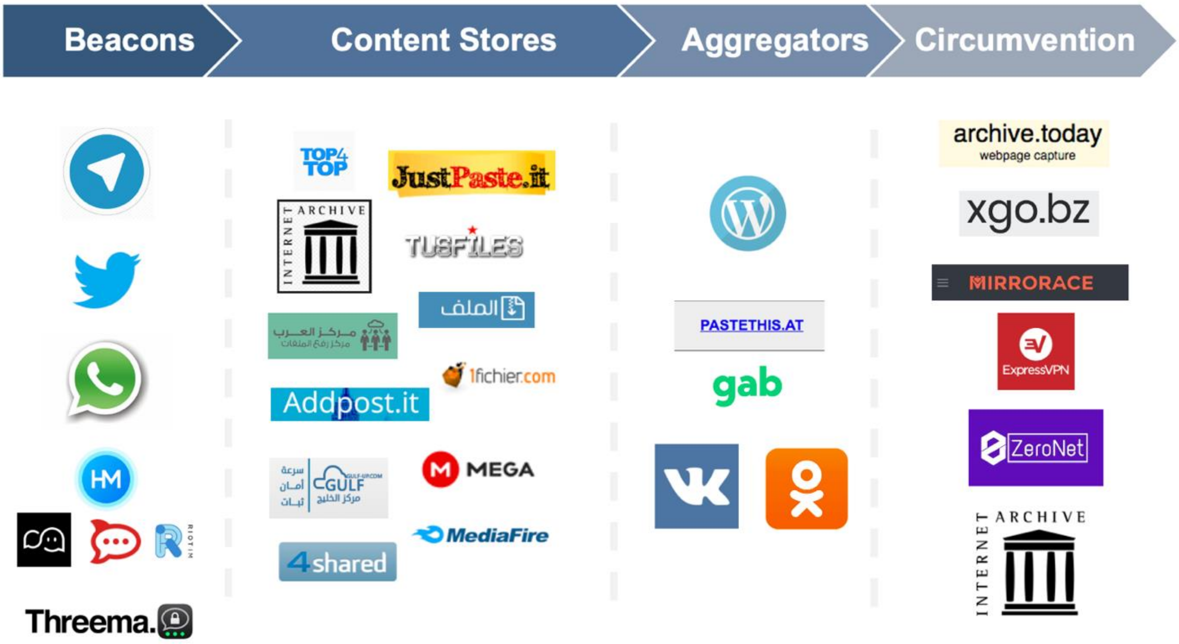


Figure 6: An illustration of how T/VE actors use a range of different platforms to disseminate content. Source: Tech Against Terrorism.

Terrorist Use of the Internet: Strategic and Operational Purposes

Terrorists and violent extremists use the internet for two main purposes: internal and external communication. External communication mostly entails the dissemination of propaganda to as wide an audience as possible, with the intention of spreading fear, encouraging attacks or inciting violence generally, and claiming attacks, as well as for recruitment. Internal communications typically involve conversations about logistics, planning and other private matters.

Strategic purposes: Gaining support and legitimacy via propaganda

Terrorist actors often instrumentalise the internet in order to legitimise and to increase support for their political objectives. T/VE groups prioritise the internet over other forms of communication given its wide potential reach, which makes online spaces attractive for the dissemination of propaganda and recruitment efforts.[16] Central to the terrorist use of the internet for strategic purposes is propaganda, which relies on facilitated communication and the manipulation of information.

Online platforms provide terrorists and violent extremists with an efficient and straightforward way in which to spread their message, raise their profile and attract support. To disseminate their content to the widest audience possible, T/VE actors often exploit multiple platforms simultaneously. To do so, terrorists rely on a multiplatform approach,[17] and on strategies that Ali Fisher has identified as a "Netwar" perspective" and "Swarmcast".[18]

- 'Netwar': Using David Ronfeldt and John Arquille's theory of 'Netwar', Fisher frames terrorist use of the internet as a networks-based form of organisation that relies on strategies characteristic of the information age. Fisher's research shows that terrorists exploit communications technologies 'to propagate awareness among the general public in the hope of mobilising it'. They do so primarily by evading law enforcement and content moderators to infiltrate, disperse and disrupt the online space.

- The 'swarmcast' model: This dissemination strategy is underpinned by an emphasis on speed, agility, and network resilience. Terrorist and violent extremist networks mount coordinated and convergent attacks, emanating from multiple axes simultaneously. Fisher shows that T/VE networks online have evolved from a centralised propaganda dissemination strategy to one executed by a diffuse and fluid network of supporters and self-appointed members who constantly upload and redistribute content. 'Official' accounts still exist, but the supporters' network ensures that content remains available and widespread. In the

[16] Propaganda is used here to comprise any content intended to generate support and enlist new recruits, or to incite individuals to act on behalf of an ideology and solicit general support. Propaganda can take different forms, from videos and audio files to posters and memes, but thematically it is usually centred around the promotion of violence. Typical propaganda content can be the footage of an attack, videos of training camps, or direct calls to violence. Propaganda can also be tailored to target a specific population, or an individual identified as a potential recruit, whether that be in propagating a sense of fear in the collective or in tailoring a propaganda strategy to one's belief. See: Gill Paul, Corner Emily, Conway Maura, Thornton Amy, Bloom Mia, and Horgan John (2017), "Terrorist use of the internet by the numbers", in *Criminology & Public Policy*, Vol. 16, Issue 1, pp.99-117; United Nations Office on Drugs and Crime – UNODC (2012), The use of the internet for terrorist purposes; Bertram Luke (2016), "Terrorism, the Internet and the Social Media Advantage", in *Journal for Deradicalization*, No. 7; and Speckhard Anne and Bodo Lorand (2018), "Fighting IS on Facebook – Breaking ISIS brand counter-narratives project", ICSCVE Research Reports

[17] See: Fisher Ali (2015), "Swarmcast: How Jihadist Networks Maintain a Persistent Online Presence" Ali Fisher, in *Perspectives on Terrorism*, Vol. 9, Issue 3; Fisher Ali, Prucha Nico, Winterbotham Emily, (2019), "Mapping the Jihadist information ecosystem: Towards the next generation of disruption capability", Global Research Network on Terrorism and Technology, Paper No. 6.

[18] Fisher (2015)

> case of jihadists, the ‘media mujahideen’ - a term used by IS itself,[19] adopts ‘genuine swarming behaviours’ online.[20]

The below graph visualises how the “swarmcast” dissemination strategy is executed in practice by T/VE actors. The graph depicts the outlinks on a violent Islamist beacon channel in 2020, collected during Tech Against Terrorism’s monitoring.

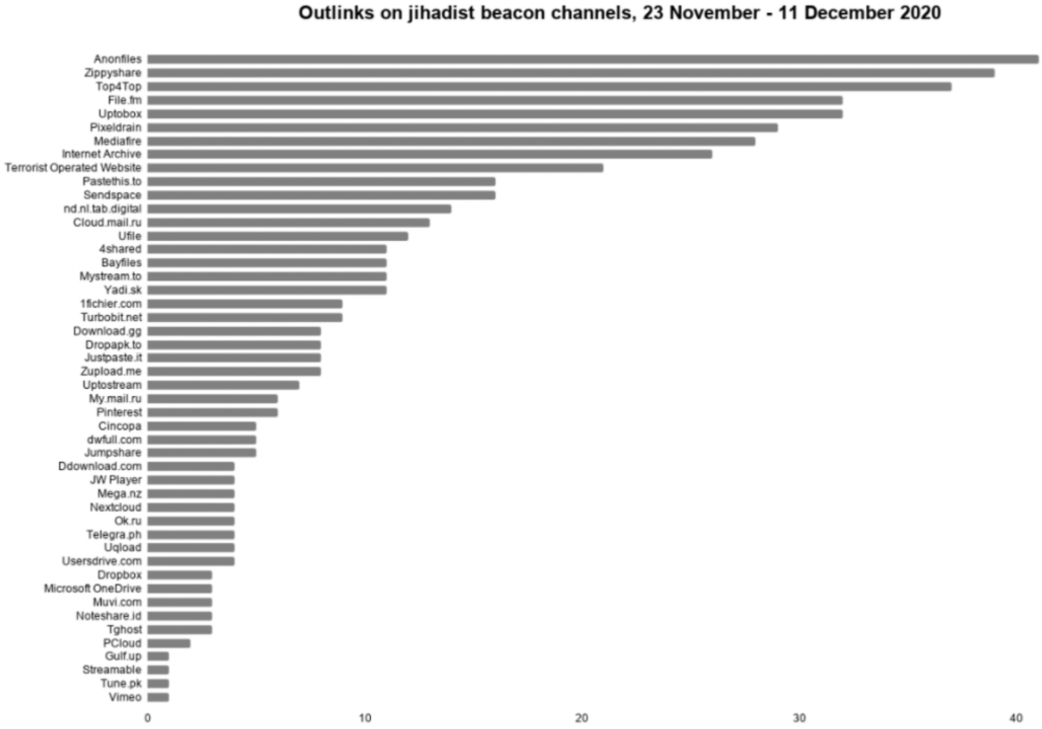


Figure 7: Graph highlighting how T/VE actors use multiple platforms simultaneously to disseminate content. Source: Terrorist Content Analytics Platform (TCAP).

[19] Winter Charlie (2017), Media Jihad: The Islamic State’s Doctrine for Information Warfare, in *The International Centre for the Study of Radicalisation and Political Violence*.

[20] Fisher (2015)

Operational purposes: Facilitated coordination and command control

Terrorists and violent extremists were quick to exploit internet technologies for operational purposes. Whether that be for fundraising – as financing is a crucial component of any terrorist organisation's survival and capacity to carry out attacks – or for planning and coordinating attacks, using open-source intelligence to inform their plans and coordinate ahead of an attack.

The key privacy and security features offered by online platforms are indispensable for the operational purposes of terrorists and violent extremists. T/VE actors have been relying in particular on end-to-end encryption, password protected websites and forums, and online peer-to-peer networks and transactions to organise themselves, raise funds and coordinate their attacks.

T/VE actors using such platforms benefit from instant reach, an almost unlimited geographic scope, as well as private and secure communications that allow minimal to no law enforcement detection.

Use of social media platforms for strategic and operational purposes

- *Strategic - Propaganda*: Use of social media platforms to publish statements and propaganda videos, as well as to share links to redirect supporters to content hosted on other platforms (including to more closed platforms and terrorist-operated websites).[21] On Instagram, the narratives shared on public pages would usually focus on less violent content – for instance memes propagating radical right ideas or content supporting extremist religious beliefs, both by omitting direct calls to violence or by using the cover of irony/humour.[22]
- *Strategic – Targeted recruitment:* Terrorists and violent extremists can use these platforms for targeted recruitment due to large audience sizes on many social media platforms.[23]
- *Strategic – Promoting a sense of community:* As for any social media user, social media are used as a place to network and create a sense of community amongst terrorists and violent extremists. Closed groups are preferred for free discussion with like-minded individuals and for organisation of offline events. [24]
- *Strategic – Pledges of allegiance*: Social media platforms have been used by 'lone wolf attackers to pledge allegiance to a terrorist group right before committing their attacks. Facebook and Instagram, for instance, have both been used by extremists in Europe to pledge allegiance to the Islamic State before carrying out their attacks.[25]

[21] Waters Gregory and Postings Robert (2018), "Spiders of the Caliphate: Mapping the Islamic State's Global Support Network on Facebook", The Counter Extremism Project.

[22] Sceckhard 2020; The Tech Against Terrorism Podcast with Fielitz Maik and Bogerts Lisa, (2020) Far-Right Violent Extremists and Meme Culture, Tech Against Terrorism.

[23] See: Davey Jacob, Hart Mackenzie, and Guerin Cecile (2020) "An online environmental scan of right-wing extremism in Canada", Institute for Strategic Dialogue; Berger J.M and Perez Heather (2016), "The Islamic State's Diminishing Returns on Twitter: How suspensions are limiting the social networks of English-speaking ISIS supporters", GW Program on Extremism;
Speckhard and Lorand (2018); and Waters and Postings (2018)

[24] See: Conway Maura, Scrivens Ryan, Macnair Logan (2019), "Right-Wing Extremists' Persistent Online Presence: History and Contemporary Trends", International Centre for Counter-Terrorism; Ekman Mattias (2018), "Anti-refugee mobilization in Social Media: The Case of Soldiers of Odin", in *Social Media + Society*, Vol. 4, No. 1.

[25] E.g.: Munir Hassan Mohammed who attempted to carry out a bombing in London on Christmas 2017 pledged allegiance to IS on Facebook. Kujtim Fejzullai who killed 4 and wounded 25 in a mass shooting attack in the Vienna in November 2020, used Instagram to publish his pledge of allegiance to IS before the attack.
See: Steinbuch Yaron (2020), Vienna terrorist Kujtim Fejzulai took selfie with AK-47, machete before attack, The New York Post.

- *Operational – Targeting opponents:* Terrorists have used social media – including Facebook,[26] Instagram, Twitter, and Omegle – to directly target their opponents. This is achieved both by hacking accounts and posting threatening messages (targeted at a specific individual or general),[27] or by propagating disinformation and hate speech online which prompts spontaneous attacks by supporters on identified targets.[28]
- *Operational – Fundraising:* The audience reach allowed by social media makes platforms a natural choice for fundraising campaigns and calls for donations . Facebook, and more recently Instagram have been used by terrorist actors to publicise fundraising campaigns.[29]

[26] On Facebook, experts have identified "coordinated raids" launched by IS supporters on Facebook, aimed at overloading targeted pages with threatening messages, divisive messages and terrorist content. See: Ayad Moustafa and Weiss Michael (2019), "Al Qaeda's Master Terrorists Are Still on Facebook and YouTube", The Daily Beast.

[27] Waters Gregory and Postings Robert (2018), Facebook", The Counter Extremism Project.

[28] Aaditya Dave (2019), "Transnational Lessons from Terrorist Sue of Social Media in South Asia", in *Global Research Network on Terrorism and Technology,* Paper No. 13

The most recent example of such use of Facebook to share disinformation is the far-right extremist and violent extremist use of the platform to spread mis and dis-information regarding the Covid-19 health crisis, often including hyperlinks to external platforms in an effort to use the platform as a gateway portal for further extremist engagement. See: Boyer Iris, Lenoir Théophile (2020) "Information Manipulations Around Covid-19: France Under Attack", Institut Montaigne and Institute for Strategic Dialogue; Institute for Strategic Dialogue and BBC (2020) Covid-19 misinformation briefing No.3 – Far-right exploitation of Covid-19

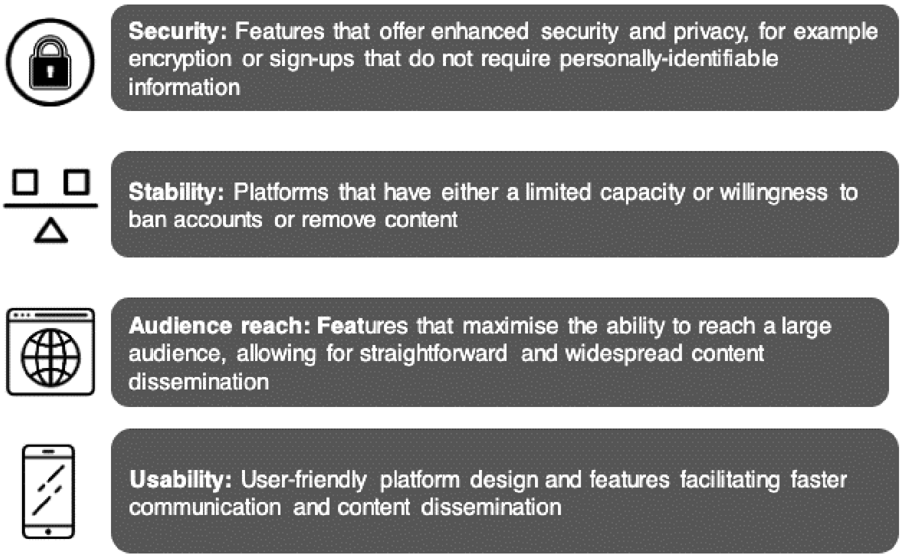
[29] Farivar Cyrus (2020), "Feds announce largest seizure of cryptocurrency connected to terrorism", NBC News; Keatinge, Keen Florence, and Izeman (2019)

What makes a platform attractive to terrorists?

T/VE actors are more likely to exploit platforms that provide offerings and features they believe will help them meet their strategic and operational objectives. Experts have shown that terrorists and violent extremists search for three main categories of features in an online platform/app: security, stability, and audience reach.   To these three categories, Tech Against Terrorism has added a fourth one: usability.[30]

Terrorists will aim to use platforms that encompass all four of these categories. But moderation efforts by tech platforms, alongside T/VE specifications for external vs. internal communications, mean that terrorists must frequently balance or prioritise the various sets of features available. A terrorist organisation may opt for an encrypted messaging platform as its 'safe haven', for example, on the basis that it cannot be de-platformed and will remain "safe" there.

But in instances where a terrorist actor or group wants their message to be amplified as far as possible, they may prioritise audience reach features over security, and set up accounts on more mainstream social media platforms.



Security: Features that offer enhanced security and privacy, for example encryption or sign-ups that do not require personally-identifiable information

Stability: Platforms that have either a limited capacity or willingness to ban accounts or remove content

Audience reach: Features that maximise the ability to reach a large audience, allowing for straightforward and widespread content dissemination

Usability: User-friendly platform design and features facilitating faster communication and content dissemination

The following table summarises some of the key features or characteristics for each of these categories, and whether they are preferred by terrorists and violent extremists for strategic or for operational purposes.

[30] Tech Against Terrorism (2019).

	Platforms used for Operational Purposes (Internal communications)	Platform used for Strategic Purposes (External communications)
Security	▪ Private chat ▪ Closed servers and forums (access granted subject to approval by the administrators or moderators) ▪ End-to-end encryption ▪ Deletion / timed destruction of messages ▪ Password-protection ▪ No phone number required upon registration ▪ Invite-only access ▪ Easy account deletion / data erasure ▪ Screenshot alerts	▪ Private groups and profiles ▪ Direct messages
Stability	▪ No moderation possible (for instance because of E2EE) ▪ Open source	▪ Low content moderation (capacity or unwillingness)
Audience reach	▪ Voice and video calls ▪ Anonymous social media features ▪	▪ Widely available and used worldwide ▪ Public groups and profiles ▪ Large size groups and broadcast lists ▪ Livestreaming ▪ Stories (video) ▪ Ability to search for groups (within the platform or through a search engine) ▪ Supports multimedia ▪ Location-based chats ▪ Scheduled messages ▪ Invite links and account IDs easily sharable
Usability	▪ Secure file-storage ▪ Important file-sharing and storage capacity ▪ Possibility to have multiple accounts ▪ Possibility to access the account from multiple devices ▪	▪ Free ▪ User-friendly interface ▪ Forwarding (ideally with no limits) ▪ Emojis, memes and stickers creation ▪ Accessible in different languages (e.g., Arabic, Spanish, Hindi, French, Portuguese)

Case study: Telegram Messenger

Telegram Messenger provides an illustrative example of how a combination of features can attract terrorists to tech platforms. Telegram offers a broad range of features that are attractive to terrorists and has been widely exploited by terrorist organisations and their supporters, including Islamic State, al-Qaeda and extensive networks of violent far-right extremists.

Telegram has been more widely exploited by terrorists and violent extremists in recent years.[31] In our analysis this is because Telegram offers a broad range of features that provide terrorists and violent extremists with an app that is easy to use, both secure and more stable than other platforms, and that allows them to reach a wider audience.

Telegram: An app of choice for terrorists	
Audience Reach	▪ Public groups and channels are searchable within Telegram and are openly accessible to all users ▪ Private groups and channels that are not searchable can be accessed via a join link, which can be promoted on public channels and third-party platforms ▪ Supergroups can include as many as 200,000 members ▪ Channels can broadcast to a theoretically unlimited number of users
Security	▪ Secret chats are protected by end-to-end encryption. Content and messages contained can only be accessed on the device of origin or destination ▪ Secret chats include a self-destruct feature ▪ Public and private groups and channels are protected by client-server/server-client encryption ▪ Telegram is founded on principles that emphasise user privacy. It states on its website that "protecting your private conversations from snooping third parties such as officials" is an essential foundation of the platform[32]
Stability	▪ Channels, and the messages and multimedia content contained within, often remain available after channel administrator accounts have been suspended ▪ Terms of service only explicitly prohibit the promotion of violence on "publicly viewable Telegram channels". This means that private channels and groups do not seem to be targeted by content moderation ▪ Telegram cloud servers are placed in several countries across the world to avoid any specific state from having sole jurisdiction. The company is based in Dubai, but says that it is "ready to relocate again" if local regulations change
Usability	▪ Large 2GB file size limit, allowing for sharing of large files including high-resolution feature-length videos ▪ Cloud-based storage with seamless sync. Messages are accessible from several devices simultaneously, and users can share an unlimited number of files. If users don't want data stored on their device, they can keep it in Telegram's cloud ▪ Multimedia history of channels and groups is viewable in a separate tab ▪ Channel feature allows administrators to control the flow of information. Only they can post, permitting for a unidirectional flow of information

[31] "Europol and Telegram take on terrorist content online', Europol Press Release, 25 November 2019, available at: https://www.europol.europa.eu/newsroom/news/europol-and-telegram-take-terrorist-propaganda-online; Rebecca Tan, 'Terrorists' love for Telegram, explained, *Vox*, 30 June 2017, available at: https://www.vox.com/world/2017/6/30/15886506/terrorism-isis-telegram-social-media-russia-pavel-durov-twitter, Will Bedingfield, 'How Telegram became a safe haven for pro-terror Nazis, *Wired*, 1 March 2020, available at: https://www.wired.co.uk/article/hope-not-hate-telegram-nazis, Jakob Guhl and Jacob Davey, 'A Safe Space to Hate: White Supremacist Mobilisation on Telegram', *Institute for Strategic Dialogue*, 26 June 2020, available at: https://www.isdglobal.org/wp-content/uploads/2020/06/A-Safe-Space-to-Hate.pdf.

[32] Telegram FAQs, available at: https://telegram.org/faq

Key trends in terrorist use of the Internet

The following section details how online platforms are exploited by terrorists and violent extremists and deals separately with the key trends observed at the time of writing.

Online resilience and adaptivity

Terrorists' tech capabilities are shaped by their need to maintain an online presence in spite of efforts by tech companies and governments to identify and arrest members and sympathisers, or to remove their content from the internet. This capability includes terrorist networks' multiplatform approach comprising the creation of applications and dedicated websites to ensure the ongoing availability of their content online. It also includes the tactics and guidance followed by terrorists to circumvent content moderation and identification by their perceived enemies, all of which constitutes terrorist operational security(OpSec).

Tactic	Description
Mirroring	Anticipating that their accounts, channels, servers or posts are likely to be taken down by platform administrators, terrorists and violent extremists sometimes create multiple identical accounts, or upload multiple copies of the same content simultaneously. The aim is to overwhelm content moderation teams by creating more accounts than they are capable of moderating. This tactic has been pioneered by Islamic State on Telegram, where it has simultaneously run as many as 20 mirror versions of its 'official' propaganda channel.
Private channels and/or servers	Terrorist and violent extremist organisations will often respond to takedowns of public groups and channels by creating private, invite-only versions. Depending on the platform, this will make it more difficult for content moderation teams to take the channel or group down, particularly when the channel name does not provide clues to its contents; some platforms do not moderate private channels at all. Share links to the channel can be shared within and outside the platform.
Content editing and repurposing	Content produced by terrorist organisations is often edited and repurposed to avoid automated takedowns, for example by blocking out branding or segmenting illegal content from that which is more admissible, such as mainstream media reporting.
Language amendments	Terrorists and violent extremists avoid keyword detection by tech platforms by amending terms and phrases that may already be on the radar of content moderation teams. They may insert spaces and underscores in the middle of key phrases, for example, or change their language entirely. Tech Against Terrorism has seen Telegram channels containing Arabic IS content, for example, change their titles to Mandarin. Another prominent example is the 'Boogaloo' movement, which adapted its title to other similar iterations such as the 'Big Luau' and the 'Big Igloo".
Rhetoric dilution	Knowing the terms of service of the platforms on which they are operating, many extremist individuals and organisations intentionally soften their rhetoric to avoid deplatforming. This is despite their rhetoric being often more overtly hateful or supportive of violence elsewhere. This is particularly the case with far-right (violent) extremists, who attempt to pose as legitimate, non-racist political commentators on mainstream platforms while posting more extreme content elsewhere.
Misrepresentation	Terrorists and violent extremists often exploit legal clauses in several countries that permit the sharing of terrorist content for journalistic or research purposes. Violent far-right extremists, for example, often share graphic content or instructional material alongside a deliberate caveat that they are sharing for 'journalistic' purposes, and that they 'do not endorse' the material being shared. This is a deliberate misrepresentation tactic intended to circumvent content moderation. Another example of this relating to TikTok is the "AntiMedia" network flagged by Tech Against Terrorism to TikTok in May 2021, a pro-IS network that intentionally

	misrepresented itself as a news organisation, rather than supporters of Islamic State, in order to avoid deplatforming.
Outlinking	By posting content via third-party platform outlinks, terrorists may evade detection by content moderation teams, particularly when the linked content would be picked up by automated detection systems if it were posted in-app. As outlined above, terrorists and violent extremists also often post multiple outlinks to the same content simultaneously, in the knowledge that the content is likely to be taken down. This tactic increases the changes that the content can be found on at least one of the outlinks.
Archiving	Web archiving services such as the Internet Archive are used by terrorists and violent extremists to create backed-up copies of content that has been uploaded to file sharing platforms. Many of these services are free and easy to use, and guarantee user anonymity. They also have terms of service that place the burden of responsibility on users.

Terrorist Operated Websites (TOWs)

Terrorist Operated Websites (TOWs) play an important role in the online terrorist ecosystem and serve to ensure the accessibility and ongoing availability of propaganda and other content online. At the time of writing, Tech Against Terrorism was tracking almost 140 domains on which a significant proportion of the content is terrorist or extremist in nature, or which Tech Against Terrorism assess to be entirely terrorist- or extremist-operated.

TOWs are websites that are run by terrorist groups or their supporters with the intended purpose of advancing the interests of a terrorist organisation or movement. Content hosted on such sites tends to be propaganda but may also include news reporting on topics unrelated to terrorism, or instructional materials such as on operational security or attack planning. TOWs serve as centralised, curated archives of content that may have been removed from social media platforms or messaging channels. URLs to these sites are often promoted in content posted on tech platforms. A list of terrorist-operated domains monitored by Tech Against Terrorism is available on request.

One example of this is Elokab, a publicly available archive of Islamic State content containing more than 90,000 pieces of propaganda material, and comprising more than 2,000 separate web pages. Unlike accounts on third-party platforms like Facebook, Twitter or Telegram, terrorists are able to control content on websites given that individual posts or pieces of content are not liable to content moderation.

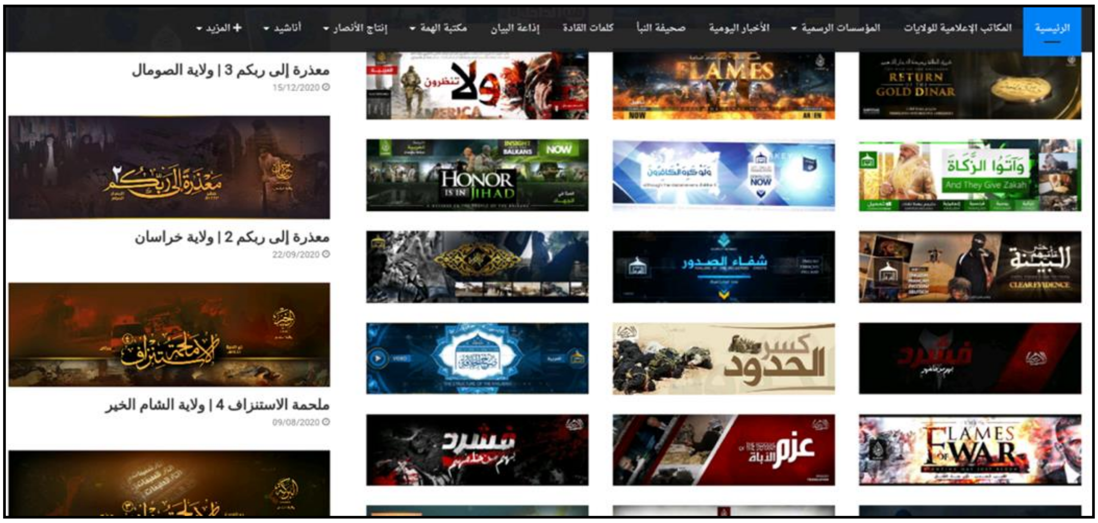


Figure 8: A screenshot of Elokab, a pro-IS propaganda website.

Terrorist Operational Security (OpSec)

Terrorists and violent extremists place great emphasis on operational security online, particularly on how to keep their communications and location data private. Dedicated opsec groups have long provided members of terrorist networks and organisations with advice on staying anonymous online, as well as producing guides on the devices, software, platforms and tactics that will best ensure their users' safety online, as well as their anonymity.[33]

Two prominent examples of online groups dedicated to providing opsec advice to terrorists are the Electronic Horizons Foundation (EHF) and OpsecGoy, each dedicated to advising violent Islamist and violent far-right networks, respectively, on operating safely and securely online.

Focus on: Electronic Horizons Foundation (EHF)

The tech knowledge-sharing arm of IS since 2016, the EHF started as an 'IP support desk' for IS supporters. It has become a well-known opsec portal in the terrorist online space in the past few years via its dedicated website. Its content is widely shared among IS supporters online. The site is registered with Netim on a .io domain and is supported by CloudFlare. A previous iteration of the site was taken down by Tucows, its previous Registrar, following reports by Tech Against Terrorism in September 2020.

EHF posts regular 'Tech News Bulletins', covering the latest news of interest to security-conscious violent Islamist extremists. The site focuses exclusively on sharing online security tips rather than terrorist content. As such, the site is an example of the difficulties facing tech platforms in moderating extremist content, as knowledge of the links between OHF and pro-IS networks would be the only identifier of the site as being terrorist-operated.

App development

Terrorists and violent extremists have also created their own apps and video games. Islamic State, for example, has developed an app to disseminate tweets more efficiently.[34] *Dawn of Glad Tidings* allowed them to use 'external Twitter accounts linked to the application to tweet statements, despite their own known Twitter accounts having been suspended'.[35] In a more recent example relating to the far-right, the German-speaking wing of the *Identitarian Movement* has developed *Patriot Beer*, which allows users to gain points for connecting with like-minded individuals, taking part in events or visiting certain cultural places. The app also permits users to find others, leading some commentators to describe it as a mix between Tinder and Pokémon Go.[36]

[33] Loedenthal Michael (2020), "Digital Resiliency and OPSEC strategies amongst clandestine networks', *Global Network on Extremism and Technology*, 10 September 2020, available at: https://gnet-research.org/2020/09/10/digital-resiliency-and-opsec-strategies-amongst-clandestine-networks/

[34] Camino Kavanagh, Madeline Carr, Francesco Bosco and Adam Hadley, "Terrorist use of the internet and cyberspace: issues and responses", in Maura Conway et al., *Terrorist use of the internet and cyberspace: issues and responses"* IOS Press (2017).

[35] "Alt-Tech: Far-right safe spaces online", Hope Not Hate, 4 November 2018, available at: https://www.hopenothate.org.uk/2018/11/04/alt-tech-far-right-safe-spaces-online/

[36] Linda Schegel, "Points, ranking and raiding the sorcerer's dungeon: top-down and bottom-up gamification of radicalisation and extremist violent', *Global Network on Extremism & Technology*, 17 February 2020, available at: https://gnet-research.org/2020/02/17/points-rankings-raiding-the-sorcerers-dungeon-top-down-and-bottom-up-gamification-of-radicalization-and-extremist-violence/.

Alt-tech platforms

Alt-tech platforms are:

- New category of platforms that offer “Free Speech” alternatives to mainstream platforms
- Widely exploited by violent ethno-nationalist and other terrorist or violent extremist groups and actors
- High risk that these platforms act as “echo chambers” for extremist communities
- Terrorist or violent extremist content is often removed less effectively due to a lack of willingness or capacity (or both) on the part of a platform’s content moderators.

Several platforms have emerged in the past few years that claim to offer an alternative to mainstream social media and other content-hosting platforms. The founders of Gab, for example, an alt-tech social media platform popular among the far-right and which has similar features to Twitter, describe themselves as being part of a ‘Free Speech Tech Alliance’ that has emerged in response to an allegedly hostile climate towards conservatives online by tech platforms perceived to have a liberal bias.[37]

Parler, another Twitter alternative, saw a huge spike in app downloads during the US election, as American conservatives who disagreed with Twitter and Facebook’s anti-disinformation efforts migrated elsewhere. Parler went offline in January 2021 after Amazon stopped hosting the network citing “violent content” hosted on the platform. Parler’s community guidelines claim to offer users a ‘nonpartisan public square’ in which ‘removing community members or member-provided content’ will be ‘kept to the absolute minimum’ by the platform’s administrators.[38] It has in recent months become a haven for the violent far-right, including US far-right group the Proud Boys and, even more recently, The National Socialist Order (NSO). The NSO claims to be the successor of Atomwaffen Division, a violent and neo-Nazi terrorist organisation that has been linked to several murders in the US.

Alt-tech platforms like Parler and Gab present unique challenges to countering terrorism and violent extremism online, as they are typically less willing to remove or otherwise moderate offending content. These platforms often gain an influx of new users following significant content moderation decisions or deplatforming on more mainstream platforms like Facebook and Twitter, especially following the suspension of former US President Donald Trump by Twitter in January 2021.[39]

Alt-tech platforms that are unwilling to effectively police terrorist or violent extremist content also increase the risk of radicalisation among their users, by facilitating an “echo chamber” in which like-minded individuals communicate without opposing views or counter-narratives. This is particularly the case with news media on alt-tech platforms, which often feature alternative outlets with overly subjective political leanings, rather than mainstream, fact-based media outlets that provide more balanced reporting.

[37] ‘Atomwaffen Division’, Southern Poverty Law Centre, available at: https://www.splcenter.org/fighting-hate/extremist-files/group/atomwaffen-division

[38] ‘Community Guidelines’, Parler.com, 7 November 2020, available at: https://legal.parler.com/documents/guidelines.pdf

[39] “Permanent suspension of @realDonaldTrump, Twitter Blog, 8 January 2021, available at: https://blog.twitter.com/en_us/topics/company/2020/suspension.html

Differences in the use of the internet across terrorist and violent extremist ideologies

The table in the compendium summarises the key differences and similarities in internet use among the two main terrorist ideologies covered in this report – Islamist extremist and the violent far-right. It also includes emerging movements with a potential for violence: QAnon, the ‘Boogaloo’ movement, and anti-Covid-19 lockdown and vaccination movements.

Violent right-wing extremist use of the internet:

Violent right-wing extremist use of the internet is characterised by:

- Increased use (and development of) alt-tech sites & chan sites
- Content moderation circumvention techniques: Use of memetic and humoristic language, “shitposting”, sanitised discourse, content format alteration
- Hostile OSINT
- Increased focus on operational security (opsec)

Right-wing extremist online networks are decentralised and dispersed. Named, cohesive terrorist organisations do not form central nodes in the online right-wing extremist eco-system in the same way that IS, and al-Qaeda do in jihadist networks. Violent far-right extremist and terrorist actors instead manifest themselves in amorphous and shifting online communities, united more by common belief systems, jokes, and community-specific slang than mass adherence or support for named international organisations.

The violent far-right operate on a broad range of online platforms. Like Islamist terrorists, they are intent on maintaining a presence on those with a large, mainstream audience in order to either radicalise or intimidate wider populations. However, an increase in the capability and willingness of larger platforms to moderate extreme right-wing content in the past few years has forced most right-wing extremists onto more niche platforms.

Features of some of these alternative platforms, such as user anonymity or audience reach, have probably also contributed to extreme right-wing migrations there. Some are explicitly promoted as ‘free speech’ alternatives to mainstream platforms, and so have attracted large numbers of right-wing extremists. Platforms that brand themselves as ‘alternative’ to mainstream ones have particularly been exploited by far-right extremists, who often are ideologically opposed to the perceived liberal/left-wing political biases of mainstream platforms like Twitter and Facebook.

We have included more detail on violent far-right use of the internet, particularly humour and “shitposting”, in the compendium.

Islamist terrorist use of the internet

Islamist terrorist use of the internet is characterised by:

- Swarmcast model of propaganda dissemination, centred around core IS and al-Qaeda propaganda outlets
- Content released on beacons is replicated on multiple file storage and paste sites, and online availability is ensured through aggregators

Violent Islamists have been prioritised by governments and technology companies in recent years, both in terms of countering violent extremism and moderating their content. They have therefore had to be more adaptive and creative online than adherents of other violent extremist ideologies. The propaganda output from the two most prominent groups, al-Qaeda and IS, is mostly structured in terms of centralised channels or accounts, which serve as beacons, and the primary source from which official content is released. Armies of supporters then actively duplicate content, or edit and repurpose it, and then reupload and reshare it, as well as creating their own.

Violent Islamist networks intentionally adopt this ‘swarmcast’ model to maintain the online presence of their accounts and channels across platforms. In the eighth episode of *Inside the*

Caliphate, an Islamic State video released in late 2018, for example, a narrator instructs the group's supporters that "if they [tech platforms] close one account", supporters should "open another three". He goes on: "if they close three, open another thirty […] with every press of a key on the keyboard, you amplify the force and reach of the explosives"

Violent Islamist organisations and their affiliated online networks use a variety of platforms in order to disseminate their propaganda and reach their target audiences. Centralised media outlets typically utilise messaging platforms or dedicated websites. Both of these mediums are often transient, and shift URLs frequently in response to deplatforming by the relevant tech providers. These outlets act as the 'beacon' for dissemination, before supporters and members save and reupload official content elsewhere, on a broad range of social media platforms, websites and messaging platforms. We have included an overview of some of the most exploited platforms by violent extremists in the compendium, based on Tech Against Terrorism's ongoing OSINT monitoring and Fisher and Pruz's research[40].

Other violent Islamist groups have been targeted less by content moderation by tech companies and international organisations in recent years. This is in part because their impact may be less far-reaching, and more focused on domestic contexts in specific countries, or they may be armed wings of more avowedly political organisations, such as Hamas or Hezbollah. These two organisations, for example, have functional and dedicated websites, that in part due to their established political status have not been removed.

Methodology for OSINT Investigations

Tech Against Terrorism was commissioned to undertake this research by the Commonwealth Secretariat. Tech Against Terrorism conducted three separate open-source intelligence (OSINT) investigations online, focussed on the following groups: Al-Shabaab, Boko Haram and Ansar al-Sunna. Tech Against Terrorism's investigations were primarily conducted with reference to mainstream social media platforms and video-sharing platforms, namely Facebook, Instagram, Twitter and YouTube. They also involved investigating these group's online presence on smaller platforms and apps where terrorist networks are most concentrated, such as Telegram and Element Messenger, as well as on terrorist-operated websites (TOWs).

All investigations were conducted using sock-puppet accounts, as well as VPNs and virtual machines to ensure operational security. Multiple relevant keyword searches were deployed across several languages, as well as reverse image searches of certain logos, and screenshots of media.

Tech Against Terrorism's investigations also drew on third-party research conducted by other organisations and researchers. Tech Against Terrorism also consulted subject matter experts (SMEs) who focus on violent Islamist groups in Sub Saharan Africa. Due to the sensitivity of their work, we do not identify these individuals in this report.

[40] Fisher Ali, Prucha Nico, Winterbotham Emily, (2019), 'Mapping the Jihadist Information Ecosystem: Towards the Next Generation of Disruption Capability', *Global Research Network on Terrorism and Technology*, Paper No. 6,

Case Study: Ansar al-Sunna

Background

Ansar al-Sunna (also known as Islamic State Mozambique province, or Ansar al-Sunna Wa Jamma, or locally as al-Shabaab) is a violent Islamist insurgency group that is mainly active in northern Mozambique, particularly Cabo Delgado province and areas along Mozambique's border with Tanzania. Ansar al-Sunna has been linked to the Islamic State terrorist group since 2019, though the extent of their relationship remains disputed. Ansar al-Sunna began as a religious organisation in 2015 and was designated along with its leader, Abu Yasir Hassan, as a global terrorist group in March 2021 by the US Department of State. [41]

When the group initially formed, its leaders sought to take advantage of Cabo Delgado's high youth unemployment, stagnant economic growth and large Muslim population for recruitment. Its members targeted poor, marginalised, and unemployed youth for recruitment, especially among the Kimwani ethnic group in Cabo Delgado.[42]

In 2017, the group began launching attacks on government and civilian targets, particularly in the Cabo Delgado region. There have been numerous reported attacks on government buildings, mosques and civilians in the subsequent years.[43] Since October 2017, the group is alleged to have killed more than 1,300 civilians and displaced 670,000 within northern Mozambique as a result of their ongoing violent insurgency.[44]

Since March 2020, the frequency of violent attacks by Ansar al-Sunna on government and civilian targets has increased further. Notably, the group captured Mocimboa da Praia in August 2020, [45] and launched a violent assault on the city of Palma in March 2021, killing more than 40 people.[46]

[41] https://www.counterextremism.com/extremists/abu-yasir-hassan ; https://www.state.gov/state-department-terrorist-designations-of-isis-affiliates-and-leaders-in-the-democratic-republic-of-the-congo-and-mozambique/

[42] https://jamestown.org/program/ansar-al-sunna-a-new-militant-islamist-group-emerges-in-mozambique/

[43] https://institute.global/policy/mozambique-conflict-and-deteriorating-security-situation

[44] https://www.amnesty.org/en/documents/afr41/3545/2021/en/

[45] https://www.lawfareblog.com/islamic-state-mozambique

[46] https://ctc.usma.edu/the-march-2021-palma-attack-and-the-evolving-jihadi-terror-threat-to-mozambique/

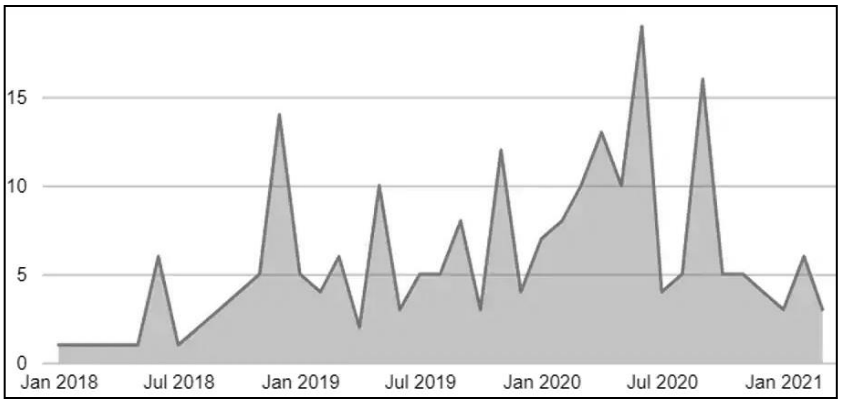
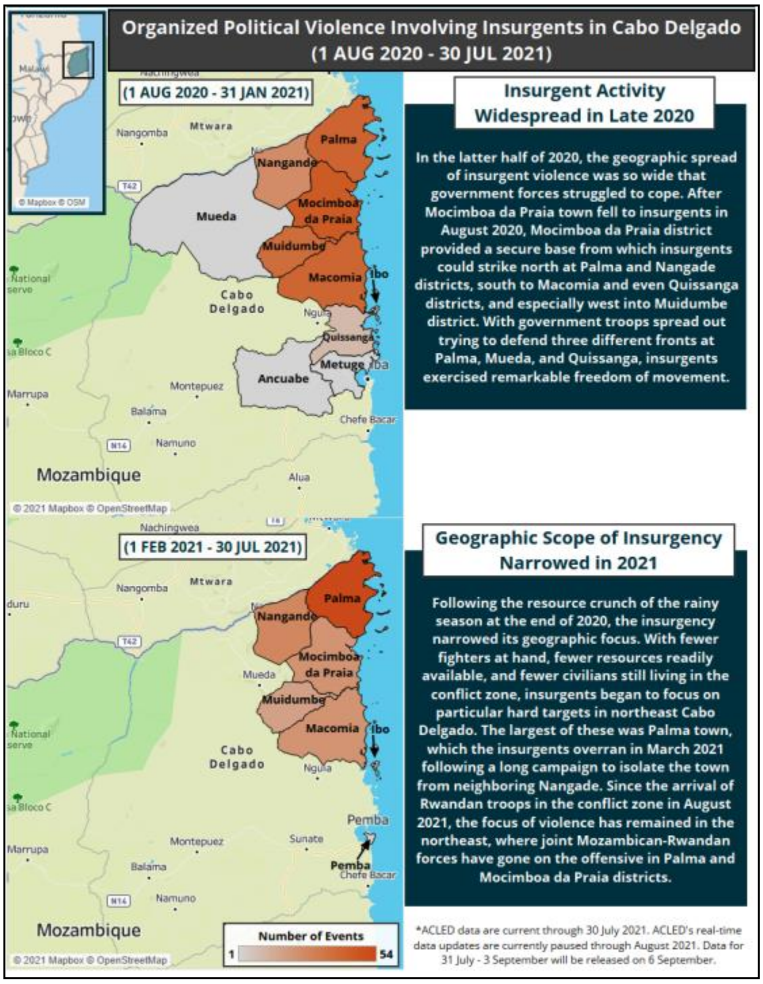


Figure 9: Reported attacks in Cabo Delgado, Mozambique between 01 January 2018 – 21 March 2021. Source: HIS Markit.[47]


Organized Political Violence Involving Insurgents in Cabo Delgado (1 AUG 2020 - 30 JUL 2021)

(1 AUG 2020 - 31 JAN 2021)

Insurgent Activity Widespread in Late 2020

In the latter half of 2020, the geographic spread of insurgent violence was so wide that government forces struggled to cope. After Mocimboa da Praia town fell to insurgents in August 2020, Mocimboa da Praia district provided a secure base from which insurgents could strike north at Palma and Nangade districts, south to Macomia and even Quissanga districts, and especially west into Muidumbe district. With government troops spread out trying to defend three different fronts at Palma, Mueda, and Quissanga, insurgents exercised remarkable freedom of movement.

(1 FEB 2021 - 30 JUL 2021)

Geographic Scope of Insurgency Narrowed in 2021

Following the resource crunch of the rainy season at the end of 2020, the insurgency narrowed its geographic focus. With fewer fighters at hand, fewer resources readily available, and fewer civilians still living in the conflict zone, insurgents began to focus on particular hard targets in northeast Cabo Delgado. The largest of these was Palma town, which the insurgents overran in March 2021 following a long campaign to isolate the town from neighboring Nangade. Since the arrival of Rwandan troops in the conflict zone in August 2021, the focus of violence has remained in the northeast, where joint Mozambican-Rwandan forces have gone on the offensive in Palma and Mocimboa da Praia districts.

*ACLED data are current through 30 July 2021. ACLED's real-time data updates are currently paused through August 2021. Data for 31 July - 3 September will be released on 6 September.

Figure 10: Maps showing areas of violence involving insurgents in Cabo Delgado, Mozambique.[48]

Affiliation with Islamic State

The extent of Ansar al-Sunna's affiliation with IS remains unclear. While official IS media entities have recognised the relationship between IS central and the insurgents in northern Mozambique,

[47] https://ihsmarkit.com/research-analysis/terrorism-mozambiques-cabo-delgado-data.html
[48] https://acleddata.com/acleddatanew/wp-content/uploads/2021/08/Cabo-Ligado-61.pdf

the insurgency was largely born out of localised domestic grievances as opposed to ideological affiliations with global violent Islamist ideologies. While Ansar al-Sunna first pledged allegiance to IS in 2018, it was only in 2019 that IS central recognised the affiliation. Since then, there has been little evidence to suggest any coordination between IS central and Mozambique, even in communications and strategies. This suggests that Ansar al-Sunna operates to a large extent separately to many other officials IS branches across the world, and that IS central exploits the group's apparent successes to project itself as having a larger global footprint than it does.

The link between the two groups was formalised after IS claimed its first attack in Mozambique in June 2019, following an attack by the insurgents against the Mozambican army in Mocimboa. IS attributed the attack to its then newly established "Central Africa Province" affiliate group.[49] IS announced the launch of the Islamic State Central Africa Province (ISCAP) in April 2019 to promote the presence of IS across Sub Saharan Africa. ISCAP is made up of fighters across the Democratic Republic of Congo and Mozambique. Although IS media entities often portray ISCAP as a unified structure, ISIS-DRC and ISIS-Mozambique are distinct groups with differing origins.

Following IS' announcement on the creation of ISCAP in 2019, the UN Security Council noted that "the online presence of ISCAP began combining footage from the Democratic Republic of the Congo, Mozambique and Somalia, an indication of coordination or attempts to unify the three theatres".[50] IS-Mozambique has been referenced in several official IS media outlets, including the group's weekly newsletter al-Naba. In an al-Naba editorial in June 2020, IS celebrated the insurgency in Mozambique and taunted western and African states for their failure to defeat the insurgents.

Amnesty International identified over 40 videos and public statements from official IS outlets over December 2019 - November 2020 that claimed credit for Ansar al-Sunna attacks.[51] However, researchers noted that not all Ansar al-Sunna attacks were claimed by official IS media.[52] The inconsistent coverage of Ansar al-Sunna's activities by official IS outlets is indicative of a more tenuous link between the two groups.

IS likely affirmed their relationship with violent Islamist insurgents in Mozambique and DRC in June 2019 following their territorial losses in Iraq and Syria two months in April 2019. IS overemphasises its global significance in order to portray itself as powerful. The group does so by leveraging loose ties with militant groups around the world.[53] Joseph Hanlon, a visiting senior fellow at the London School of Economics noted on the situation in northern Mozambique that: "This is a domestic insurgency based on domestic grievances… There are loose ties, but the insurgents have not seceded control to IS. This is not an Islamic jihad."[54]

Ansar al-Sunna's online activities

Through open-source intelligence investigations and consultations with several subject matter experts, Tech Against Terrorism identified that Ansar al-Sunna does not have a significant online presence. This is due to several reasons, including primarily:

- The group is secretive and has sought to conceal their operational and strategic activities.
- The group and the wider insurgency are highly localised, and objectives are tied to their areas of operation. Given low internet access in the northern regions of Mozambique where the group is active, Ansar al-Sunna would have few motivations for disseminating online propaganda for strategic purposes.

[49] http://www.islamedianalysis.info/islamic-state-arrival-in-mozambique-further-complicates-cabo-delgado-violence/?cn-reloaded=1&cn-reloaded=1
[50] https://clubofmozambique.com/news/un-says-insurgent-propaganda-combines-footage-from-mozambique-drc-and-somalia-carta-151858/
[51] https://clubofmozambique.com/news/un-says-insurgent-propaganda-combines-footage-from-mozambique-drc-and-somalia-carta-151858/
[52] https://www.amnesty.org/en/wp-content/uploads/2021/05/AFR4135452021ENGLISH.pdf
[53] https://www.nytimes.com/2021/03/30/world/africa/isis-mozambique-attack.html
[54] https://www.nytimes.com/2021/03/30/world/africa/isis-mozambique-attack.html

- Tech Against Terrorism's OSINT team deployed key-word searches in Kiswahili, Arabic and Portuguese for locally produced Ansar al-Sunna content across multiple social media platforms such as Facebook, Instagram, video sharing platforms like YouTube and messaging services such as Telegram.

Tech Against Terrorism found almost no indications that the group seeks to exploit online spaces for strategic or operational purposes in a coordinated way. Most branded content relating to Ansar al-Sunna since 2019 has been released by official Islamic State media channels. Even so, official IS media coverage of Mozambique is infrequent and inconsistent. The infrequent and relatively low volume of references to Cabo Delgado by IS central media entities suggests a weak link between the two groups, possibly due to an absence of regular communications. Additionally, researchers have noted there has been no sustained production of propaganda material by the insurgents themselves, apart from a handful of clips in May 2020.[55] There are several factors that explain the absence of a significant online presence.

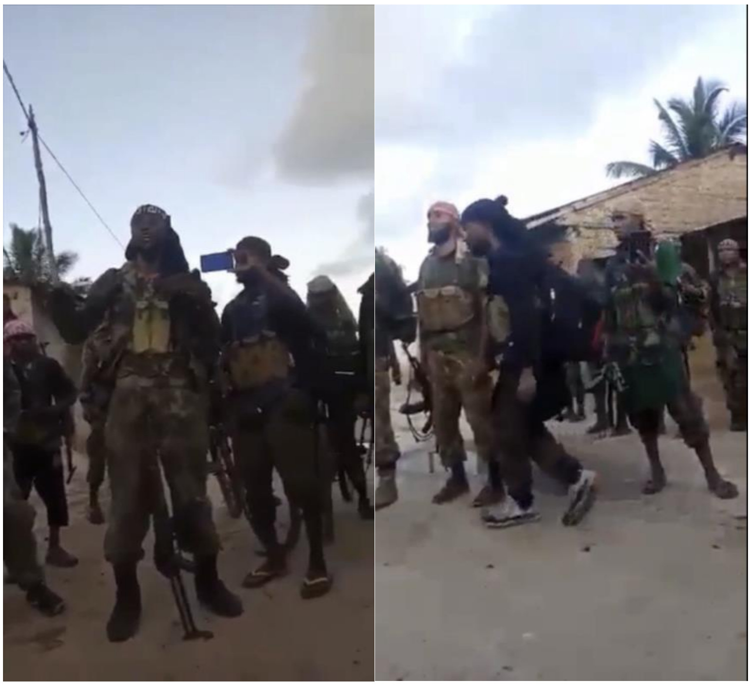


Figure 11: Screenshots of a video from April 2020 in Muidumbe depicting IS Mozambique fighters. In the left photo is Bonomade Machude Omar, who was designated as a terrorist by the US State Department in August 2021[56]

[55] https://www.caboligado.com/monthly-reports/cabo-ligado-monthly-july-2021

[56] https://www.federalregister.gov/documents/2021/08/12/2021-17277/designation-of-bonomade-machude-omar-as-a-specially-designated-global-terrorist

Figure 12: Screenshot of a video published in March 2020; depicting IS Mozambique fighters making a statement.[57].

It is likely that poor infrastructure impedes maintenance of a significant online presence. A consistent dissemination of propaganda requires stable internet access, which is limited in the northern regions of Mozambique where Ansar al-Sunna is most active. Ansar al-Sunna fighters have frequently targeted what little communications infrastructure there is in northern Mozambique, possibly so as to prevent or dramatically curtail communications.[58]

Tech Against Terrorism consulted several subject-matter-experts who argued that Ansar al-Sunna are highly secretive about their operational activities, and thus have released very few local, non-IS affiliated pieces of propaganda since 2019. Furthermore, the little content they have published has been inconsistent in volume and frequency, making it hard to track the group's output methods.

[57] https://twitter.com/emorier/status/1243082181100163072

[58] https://abcnews.go.com/International/wireStory/high-flying-balloons-boost-northern-mozambiques-internet-70660495

Figure 13: A Telegram channel that aggregates conflict-related news uploaded a video purporting to show IS fighters in Cabo Delgado, Mozambique in July 2021. The channel noted the footage was "rare" and originated from local sources. Source: ConflictNews.

In July 2021, analysts at ACLED Data said in a report that since 2019, IS-supporter groups have targeted audiences in East Africa on platforms such as Facebook, Instagram, and YouTube, and closed messaging ones such as WhatsApp, Telegram, and Element. ACLED noted multiple pro-IS Facebook accounts that mainly post in Kiswahili, Luganda and Somali in support of IS and its affiliates in northern Mozambique.[59] ACLED researchers also noted that there is a Kiswahili language podcast, distributed weekly through Facebook and other social media platforms, based on the weekly IS incident report al-Naba. It is not clear if this is supported directly by IS.[60]

These unofficial, supporter accounts post content in local languages and likely supplement the absence of official content from Ansar al-Sunna. Researchers noted that the accounts would not post overt or explicit pro-IS content, but instead spread narratives supporting certain radical clerics, or praising the insurgent's military successes. Tech Against Terrorism also learned from subject matter experts that supporter networks of Ansar al-Sunna and other violent Islamist groups often publish translated official IS content in Kiswahili and other local languages, though this content would not originate from Ansar al-Sunna itself.

[59] https://www.caboligado.com/monthly-reports/cabo-ligado-monthly-july-2021

[60] https://www.caboligado.com/monthly-reports/cabo-ligado-monthly-july-2021

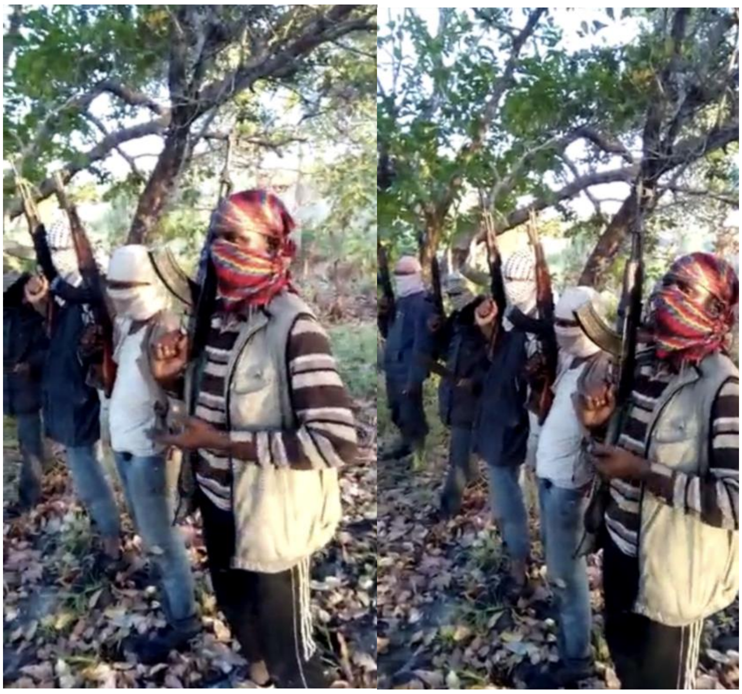


Figure 14: Screenshots of a video from January 2018 depicting IS Mozambique fighters. The precise location is unclear.

Examples of official Islamic State media content on IS Mozambique:

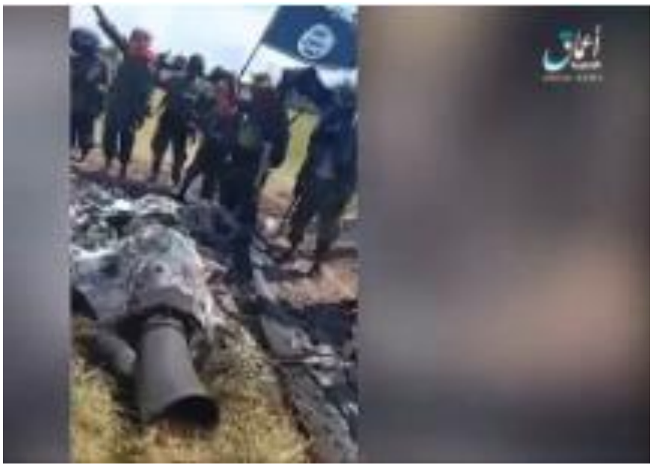


Figure 15: A screenshot of an official Islamic State propaganda video purporting to show ISCAP fighters in Quissanga, Mozambique standing near a downed military helicopter in April 2020.

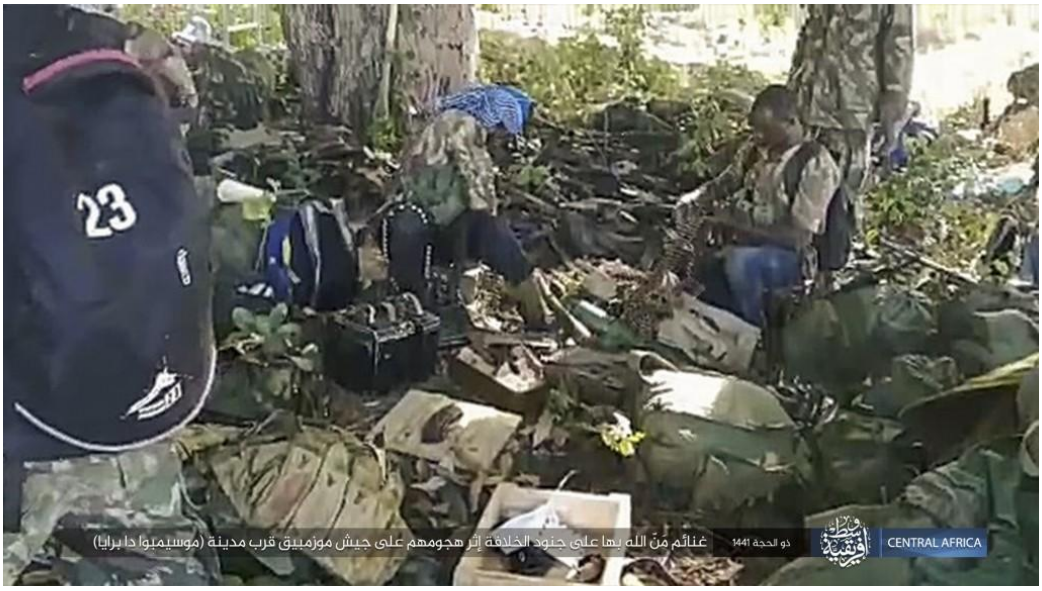

CENTRAL AFRICA
ذو الحجة 1441
غنائم مَنّ الله بها على جنود الخلافة إثر هجومهم على جيش موزمبيق قرب مدينة (موسيمبوا دا برايا)

Figure 16: Islamic State’s official Amaq news agency released images of ISCAP fighters in August 2020 following clashes in Mocimba da Praia, Mozambique.

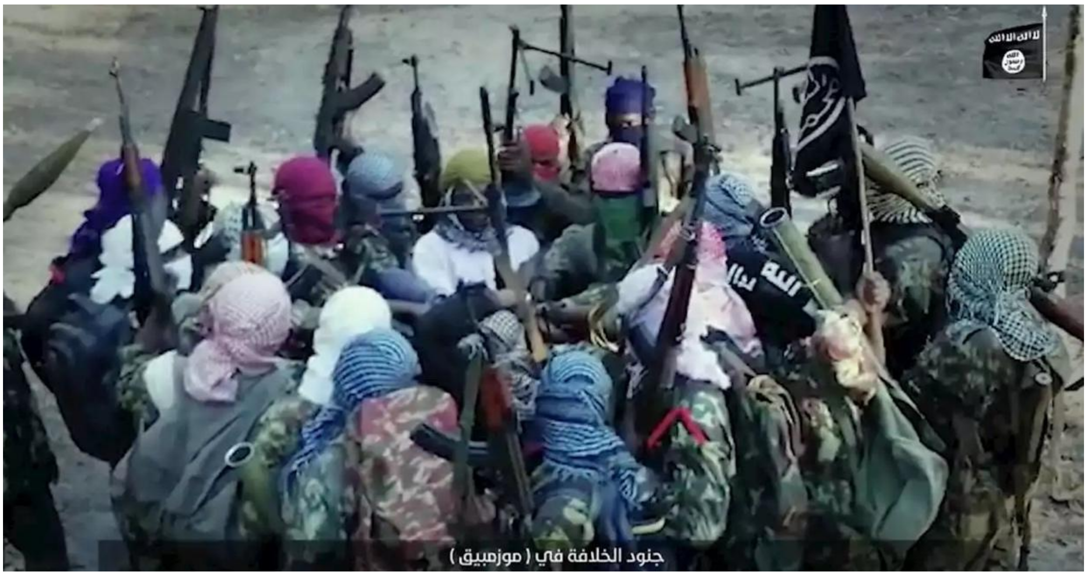


Figure 17: Official Islamic State media channels announced the formation of ISCAP in June 2015 as part of their propaganda campaign “The Best Outcome is for the Pious.” The photos depict fighters pledging allegiance to IS in Mozambique and the Democratic Republic of the Congo.

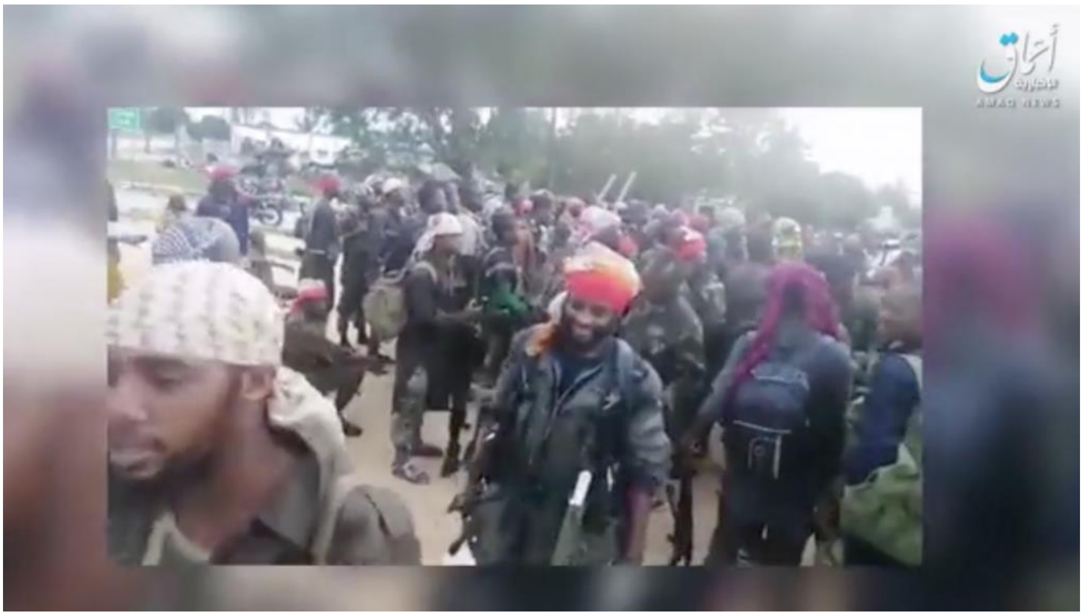


Figure 18: Screenshot of an official Islamic State video released via its news agency Amaq, purportedly showing its fighters inside Palma, Mozambique.61

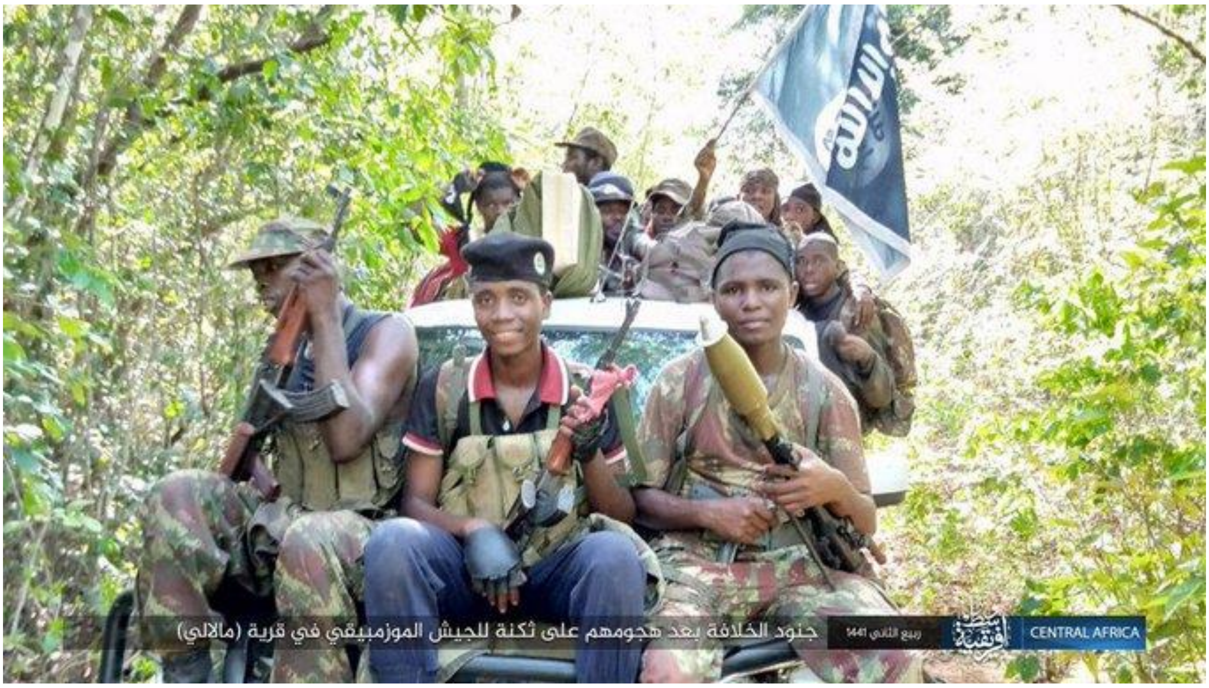

جنود الخلافة بعد هجومهم على ثكنة للجيش الموزمبيقي في قرية (مالالي)
ربيع الثاني 1441
CENTRAL AFRICA

Figure 19: Official Islamic State media agency Amaq releases a photo of fighters in Mozambique, after insurgents attack a military barracks in Mocimboa De Praia in December 2019.62

[61] https://www.longwarjournal.org/archives/2021/03/islamic-state-claims-capture-of-coastal-city-in-mozambique.php

[62] https://threadreaderapp.com/thread/1203693168170016769.html

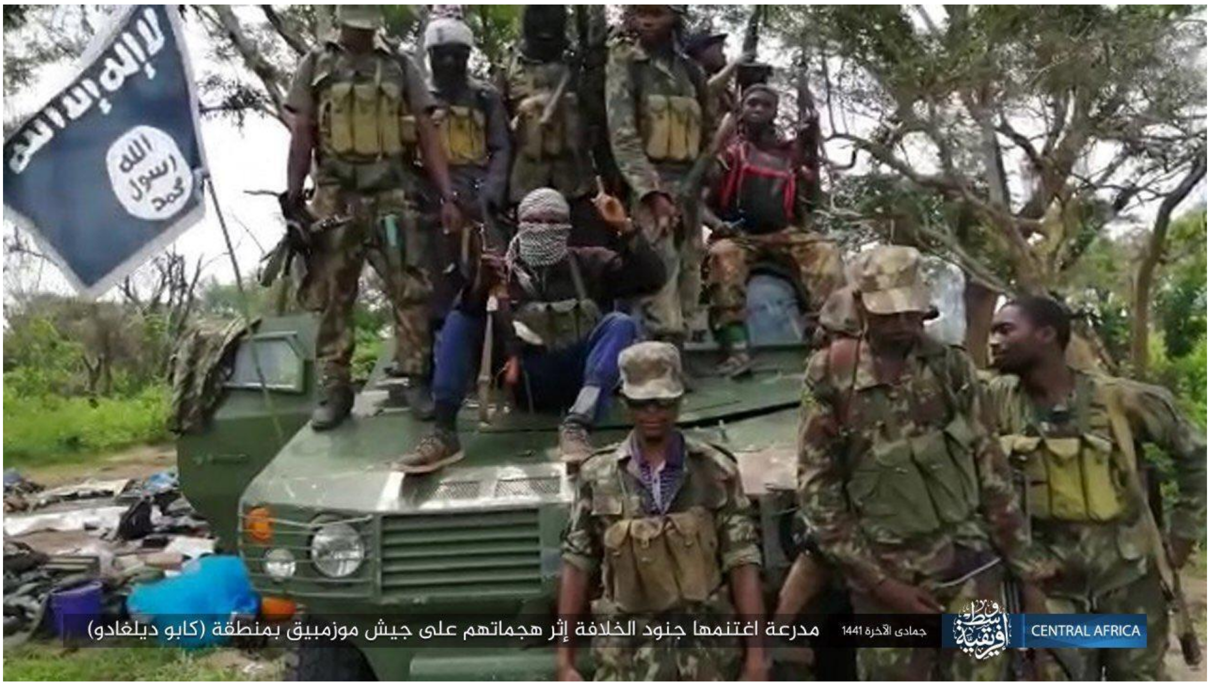

لا إله إلا الله
الله
رسول
محمد
مدرعة اغتنمها جنود الخلافة إثر هجماتهم على جيش موزمبيق بمنطقة (كابو ديلغادو)
جمادى الآخرة 1441
ولاية وسط إفريقية
CENTRAL AFRICA

Figure 20: Official Islamic State media agency Amaq publishes photos of fighters in Cabo Delgado, Mozambique via its official channels in February 2020.[63]

[63] https://twitter.com/SimNasr/status/1223678710630567936/photo/1

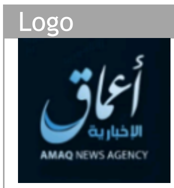
Islamic State Media

In general, official IS content is disseminated across the internet in recurring patterns. IS exploit platforms such as Telegram and Element that are used as the first access points to propaganda content. The content is usually then collated and posted on file-sharing sites, such as Archive.org or Justpaste.it, before it is then disseminated across other online spaces.

IS is also dependent on multiple unofficial, supporter-run media entities that disseminate content praising the group and attempt to spread the group's ideological message across the internet. IS supporter-run online media and news entities are created on a continuous basis and are often adept at avoiding content moderation efforts by tech platforms and law enforcement.

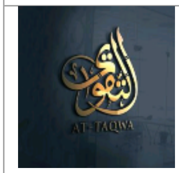
Tech Against Terrorism has compiled information on the most prominent official and unofficial IS media entities active online.

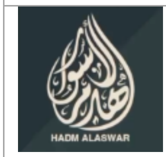
Official Islamic State Media Entities

Logo	Name	Description	Name in Arabic
	Amaq News Agency	Official IS news agency. Often the first outlet to publish claims of responsibility for attacks	وكالة أعماق الإخبارية
	Al-Bayan Radio	Islamic State's radio station	إذاعة البيان
	Al-Furqan Foundation	An official IS propaganda outlet that specialises in leadership messages	مؤسسة الفرقان
	Nashir News Agency	A propaganda outlet used to aggregate and publish official Islamic State content online	وكالة ناشر نيوز

Unofficial pro-Islamic State Media

Logo	Name	Description	Name in Arabic
	Ahlut-Tawhid	Pro-IS media outlet	أهل التوحيد
	Ajnad Foundation	An official IS propaganda outlet that specialises in audio content, particularly nasheeds	مؤسسة اجناد
	Al-Battar Foundation	Pro-IS media outlet	مؤسسة البتار إالعالمية

	Al-Dawat	Pro-IS propaganda outlet. Generally published supporter generated content, particularly posters.	الدعوات
	Al-Furat Media Foundation	A semi-official IS media outlet.	مركز الفرات لإلعالم
	Caliphate News	Pro-IS media outlet	أخبار الخالفة
	Al-Muhajireen Foundation	Pro-IS media outlet	المهاجرين مؤسسة
	Al-Mutarjim Foundation	A pro-IS propaganda outlet specialising in translating official IS content	مؤسسة ال ُمترجم
	Al-Qitaal Media Center	A pro-IS Hindi language propaganda outlet	لإعلام القتال مركز
	Al-Sumud Media Foundation	A pro-IS media outlet	الإعلامية مؤسسة صمود
	Al-Taqwa Foundation	A pro-IS media outlet.	مؤسسة التقوى
	Arrukn Media Centre	A pro-IS media outlet that claims to be based in Arakan state, Myanmar. Not recognised by IS central as an official affiliate.	مركز اركن الإعلامي
	Asdaa Foundation	A pro-IS media outlet specialising in nasheeds	مؤسسة أصداء

	Ash-Shaff Media Foundation	An English-language pro-IS media outlet	
	Hadm al-Aswar	Pro-IS media outlet. Releases posters in English threatening the West.	هدم الأسوار
	I'tisaam Media Foundation	Al-Itisaam Media is a propaganda outlet affiliated with Islamic State. Al-Itisaam Media emerged in 2013 after Islamic State in Iraq became the Islamic State in Iraq and the Levant.	مؤسسة الاعتصام
	Muntasir Media Foundation	A pro-IS propaganda group specialising in Spanish-language content.	مؤسسة منتصرالإعالمي
	Sawt al-Hind	A pro-IS propaganda group based in South Asia.	صوت الهن
	Sunni Shield Foundation	An Arabic-language pro-IS propaganda outlet.	مؤسسة الدرع السني
	War and Media	An Arabic and English-language pro-IS media group	حرب واعلام
	Urdu Nashir	Urdu-language pro-IS propaganda outlet	اردو ناشر

Case Study: Boko Haram

Background

Boko Haram – also known as Jama'at Ahl al-Sunna lil-Da'wah wal-Jihad (English translation Group of the People of Sunnah for Preaching and Jihad) – is a violent Salafist-jihadist terrorist insurgent group based primarily in Nigeria. The group aims to replace the secular Nigerian state with an Islamic state that complies with a strict interpretation of Sharia law. In recent years, Boko Haram has been fragmented by internal rivalries between the various leaders and factions that compose the main organisation.[64]

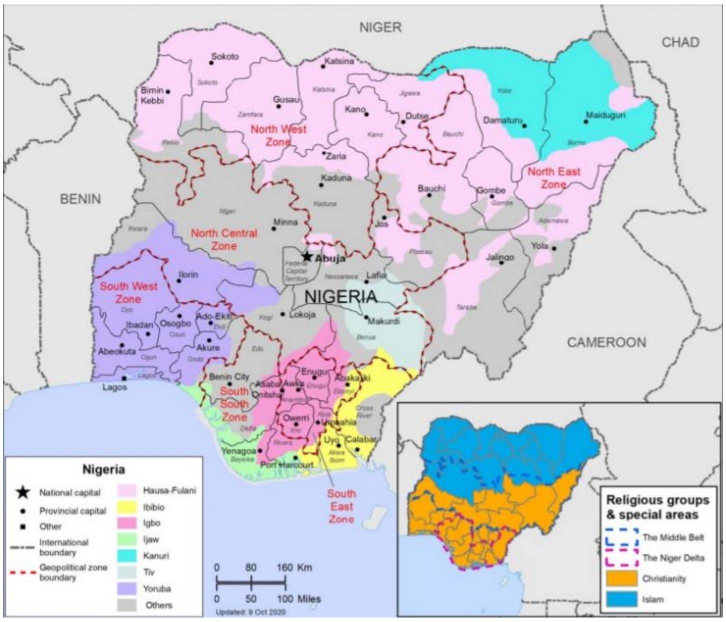


Figure 21: Map depicting ethnic, religious concentration in Nigeria. Source: UK Department of Foreign Trade and Affairs.[65]

Boko Haram has been active since 2002, particularly in northern Nigeria, and was founded by Muhammad Yusuf. who died in police custody in July 2009.[66] In 2009 the group began launching violent attacks against government military forces as well as indiscriminate attacks against civilians and humanitarian workers. The group has also conducted countless terrorist attacks on religious and political groups as well as on local police forces.[67] The group's international prominence rose

[64] https://appliednetsci.springeropen.com/articles/10.1007/s41109-020-00264-4

[65] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1003788/NGA_-_Islamist_extremist_groups_in_North_East_Nigeria_-_CPIN_-_v3.0__FINAL_Gov_UK_.pdf

[66] https://www.bbc.com/news/world-africa-57207296

[67] https://www.cfr.org/global-conflict-tracker/conflict/boko-haram-nigeria

after they kidnapped over 200 schoolgirls in April 2014. According to the United Nations Development Programme, Boko Haram's insurgency in Nigeria has killed an estimated 350,000 people since 2009 and has displaced up to 3 million civilians around the region.[68]

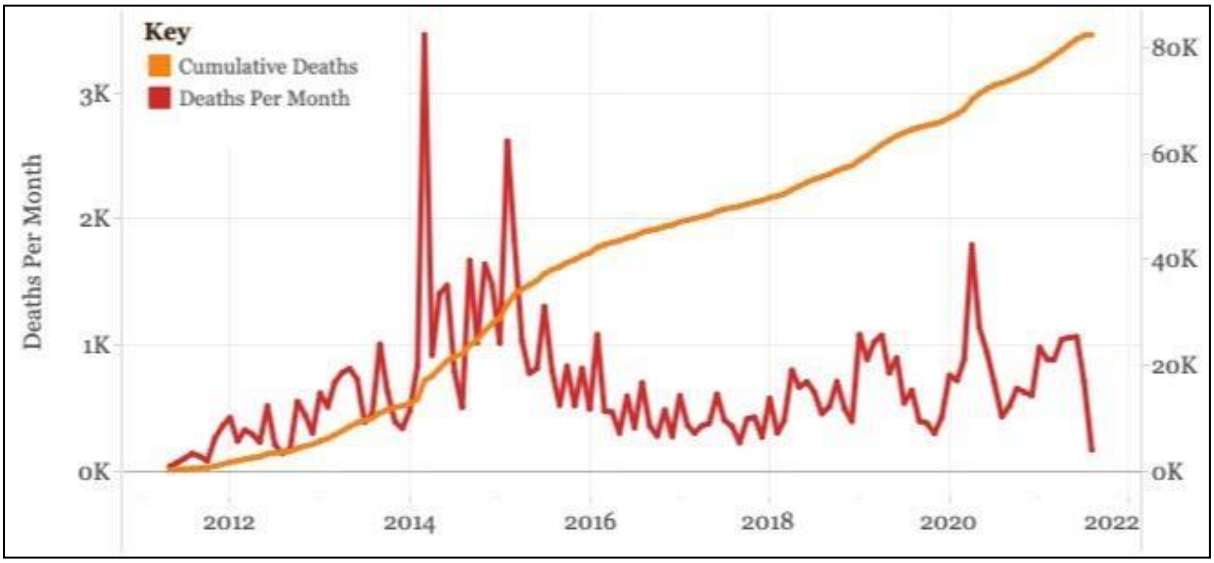


Figure 22: Deaths as a result of the insurgency in Nigeria. Source: Council for Foreign Relations.69

While Boko Haram was originally loosely affiliated with al-Qaeda,[70] the group officially pledged allegiance to the Islamic State (IS) terrorist group in March 2015, in a video message by its then leader, Abubakar Shekau. Boko Haram fragmented in 2016 following tactical disagreements with IS, thus leading to the formation of Islamic State West Africa Province (ISWAP).[71]

Until recently, Boko Haram was split between a faction led by Shekau that controlled parts of Borno State and the Cameroon-Nigeria border, and another faction led by Abu Mus'ab al-Barnawi,[72] that was mainly active in the islands of Lake Chad, West of Maiduguri and along the Niger border.[73] However, Boko Haram's leader Shekau died during clashes with rival ISWAP fighters in May 2021.[74] Since then, ISWAP has sought to consolidate its power.[75]

While the group is mostly active in Nigeria, Boko Haram also has a presence in Cameroon, Chad and Niger. The Nigerian military—with assistance from the governments of Benin, Cameroon, Chad, and Niger—has successfully pushed Boko Haram out of several provinces in north-eastern Nigeria in recent years. However, the group retains a presence in some territories and continues to launch attacks and abduct civilians, mostly women and children.[76]

[68] https://www.reuters.com/world/africa/northeast-nigeria-insurgency-has-killed-almost-350000-un-2021-06-24/

[69] https://www.cfr.org/nigeria/nigeria-security-tracker/p29483

[70] https://www.lawfareblog.com/boko-harams-al-qaeda-affiliation-response-five-myths-about-boko-haram

[71] https://www.dw.com/en/boko-haram-leader-is-dead-jihadist-rivals-claim/a-57795611

[72] http://real.mtak.hu/125093/1/Sinko-BelugyiSzemle2021.eviSPEC1.szam123-141.pdf

[73] https://appliednetsci.springeropen.com/articles/10.1007/s41109-020-00264-4

[74] https://www.bbc.co.uk/news/world-africa-57378493

[75] https://www.reuters.com/world/africa/boko-haram-fighters-pledge-islamic-state-video-worrying-observers-2021-06-27/

[76] https://www.cfr.org/global-conflict-tracker/conflict/boko-haram-nigeria

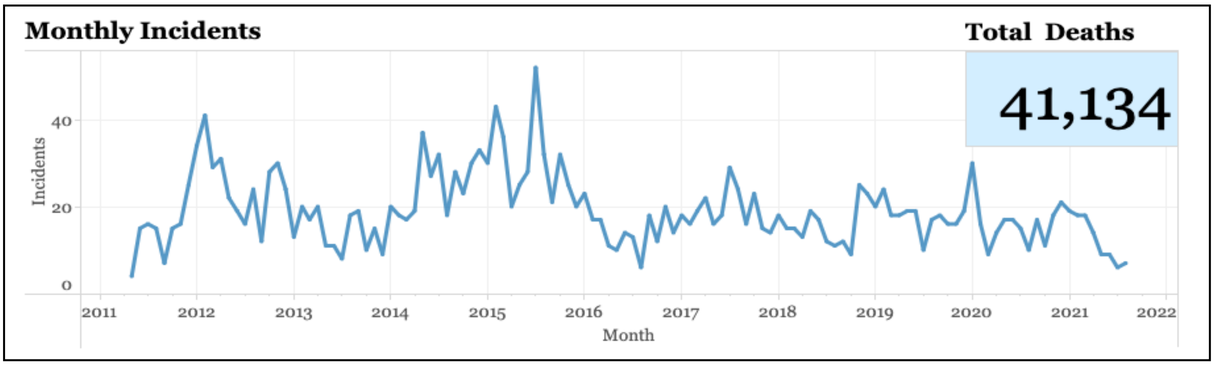


Figure 23: Monthly security incidents involving Boko Haram. Source: Council for Foreign Relations.77

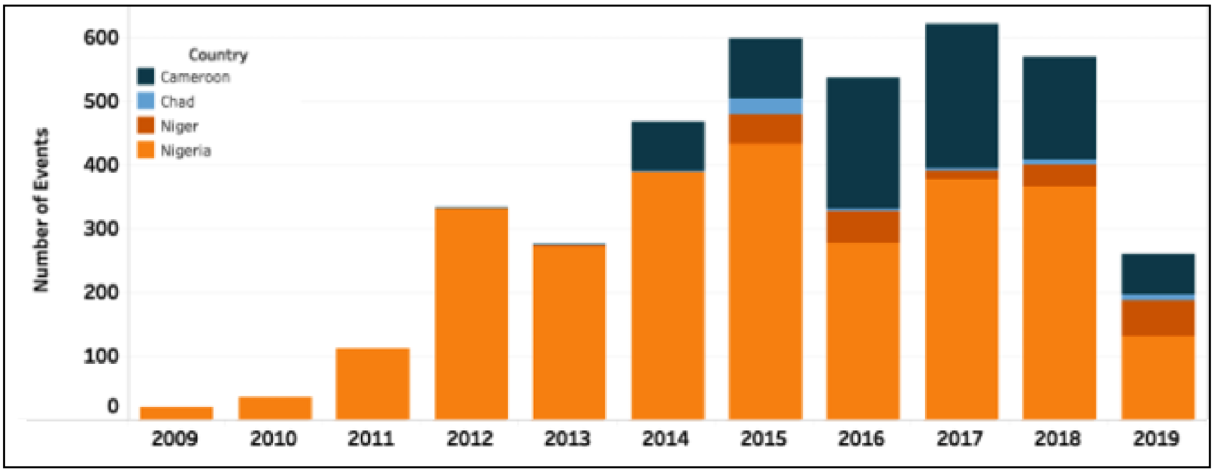


Figure 24: Reported violent events involving Boko Haram between 2009-2019 in Nigeria, Niger, Chad and Cameroon. Source: ACLED Data.78

Boko Haram's online activities

Despite many violent Islamist groups having sophisticated online information operations, Boko Haram has not had a particularly strong online presence. While Boko Haram has attempted to exploit the internet in a number of ways since 2009, overall, it has not succeeded in establishing a stable online presence. Furthermore, Boko Haram's dissemination of propaganda content has been inconsistent and at times erratic, according to some reports.[79] Tech Against Terrorism did not find evidence of Boko Haram exploiting social media platforms in open-source intelligence investigations. Tech Against Terrorism's team deployed several key-word searches across mainstream social media platforms such as Facebook, Instagram and Twitter, as well as video-streaming platforms such as YouTube. Searches were conducted in Arabic, English, and Hausa.

The group first established an online presence in 2011, however it no longer publicly disseminates content. When the group were active online, researchers noted that Boko Haram's few early videos were of low quality and were disseminated haphazardly.[80] According to a paper by the University of Swansea: "although the group does not currently have official online account(s), there are several

[77] https://www.cfr.org/nigeria/nigeria-security-tracker/p29483

[78] https://acleddata.com/2019/05/20/no-home-field-advantage-the-expansion-of-boko-harams-activity-outside-of-nigeria-in-2019/

[79] https://www.ecoi.net/en/file/local/1426834/1226_1521122538_war20.pdf

[80] https://www.nytimes.com/interactive/2015/06/11/world/africa/boko-haram-isis-propaganda-video-nigeria.html

channels hosting its contents as it leverages messaging apps to coordinate with members and with ISIS."[81]

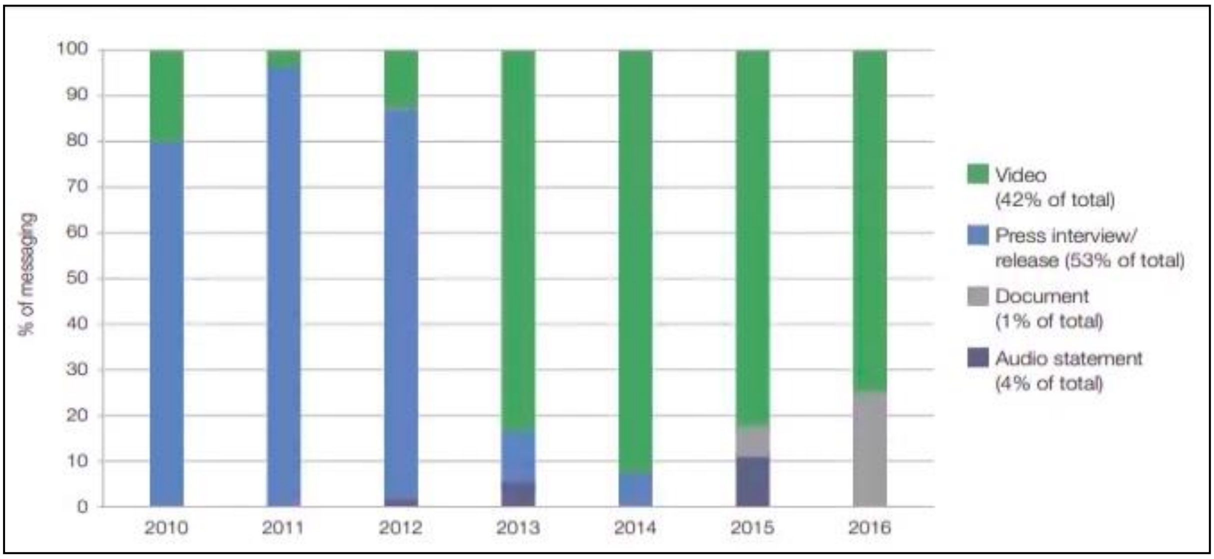


Figure 25: Breakdown of Boko Haram messaging, by type and year over 2010-2016.82

According to a report by the UNDP in 2020, Boko Haram's use of social media has been influenced by growing internet access in Nigeria in recent years.[83] From the start of Boko Haram's insurgency, internet access in Nigeria has expanded significantly. Internet usage tripled between 2012-2015, with internet penetration at around 51.4 percent of the population in 2021.[84] Despite the increase in internet access in Nigeria, it remains limited in some rural areas.

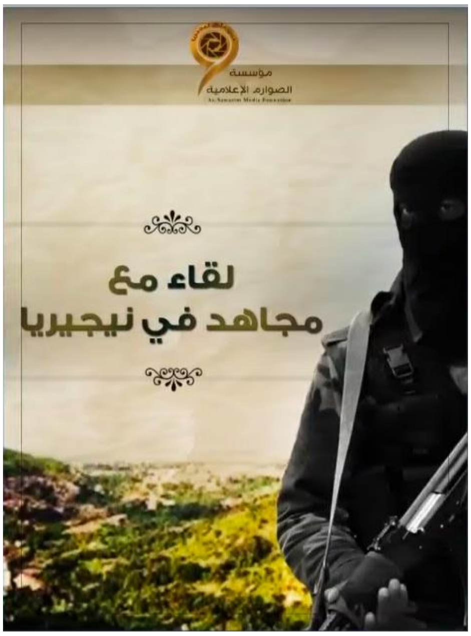


Figure 26: Boko Haram propaganda available at archive.org.

[81] https://www.swansea.ac.uk/media/TASM-Abstracts-and-Key-Messages.pdf
[82] https://www.ecoi.net/en/file/local/1426834/1226_1521122538_war20.pdf
[83] https://www.africa.undp.org/content/dam/rba/docs/Reports/UNDP-RAND-Social-Media-Africa-Research-Report_final_3%20Oct.pdf
[84] https://www.statista.com/statistics/484918/internet-user-reach-nigeria/

Much like other terrorist groups and violent extremist entities, Boko Haram has exploited social media primarily to share propaganda, attract recruits and coordinate its activities. While the group previously favoured more traditional forms of media such as audio cassettes, leaflets, open air lectures, since 2015 it has shifted to using platforms such as YouTube, Twitter and Facebook to disseminate its message.[85] Before moving to online media, Boko Haram attempted to establish a newspaper and distributed audio cassettes with recorded messages from its leaders Mohammed Yusuf and Abubakar Shekau.[86] The group's messaging largely aims to promote their successful attacks, remind supporters it holds territory, reinforce its ideology, and radicalize recruit prospective members.[87]

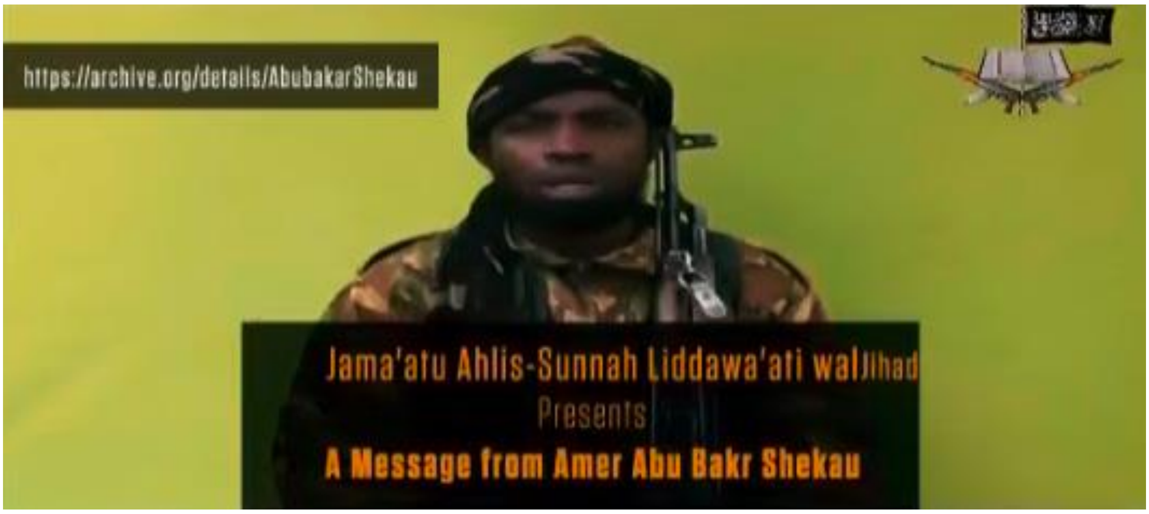

https://archive.org/details/AbubakarShekau
Jama'atu Ahlis-Sunnah Liddawa'ati walJihad
Presents
A Message from Amer Abu Bakr Shekau

Figure 27: Boko Haram propaganda featuring the groups previous leader, Abubakr Shekau on archive.org.

In 2017, the Institute for Security Studies (ISS) reported that between 22 January and 8 March 2015, in the lead-up to Boko Haram's declaration of allegiance to IS, four Twitter accounts operating under the name "al-Urwah al-Wuthqa" (English translation: The Indissoluble Link) posted Boko Haram content.[88] Each account was suspended for violating user policies in March 2015, and had gained around 4,000 in a few days. The establishment of an official Boko Haram media entity on Twitter was indicative of the Islamic State's influence over the group, argued the ISS

The Twitter accounts distributed updates on the group's operations, photos from front lines and links to video messages, tweeting in a combination of primarily Arabic and English. The final tweet published the audio speech of Boko Haram's leader pledging allegiance to IS, leading many to speculate that the al-Urwah al-Wuthqa media wing was created specifically to facilitate the admission of Boko Haram into IS. Since the suspension of that account, the al-Urwah al-Wuthqa media wing no longer exists, and Boko Haram has not attempted to re-join Twitter.[89]

[85] https://www.africa.undp.org/content/dam/rba/docs/Reports/UNDP-RAND-Social-Media-Africa-Research-Report_final_3%2520Oct.pdf&sa=D&source=editors&ust=1630936971453000&usg=AOvVaw1ysuE2qUsGP0jOurnX98Yk

[86] https://cisac.fsi.stanford.edu/mappingmilitants/profiles/boko-haram

[87] https://smallwarsjournal.com/jrnl/art/evolving-threat-us-national-security-produced-islamic-terrorist-organizations-north-africa

[88] https://www.ecoi.net/en/file/local/1426834/1226_1521122538_war20.pdf

[89] https://www.ecoi.net/en/file/local/1426834/1226_1521122538_war20.pdf

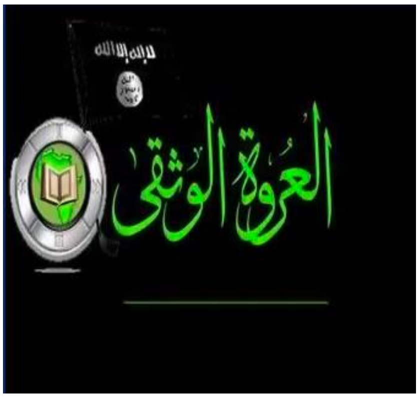


Figure 28: Logo for Boko Haram’s media wing Al-Urwa Al-Wuthqa.[90]

In November 2018, BBC Monitoring reported that Boko Haram had begun posting official content via a Telegram channel named “Attibyan”.[91] The output of the channel included a video featuring the group's leader Abu Bakr Shekau, who sought to prove that he was alive and well, following rumours about his possible ill health and death. The BBC noted that the new Boko Haram Telegram channel had copied IS’s branding in an identical manner.

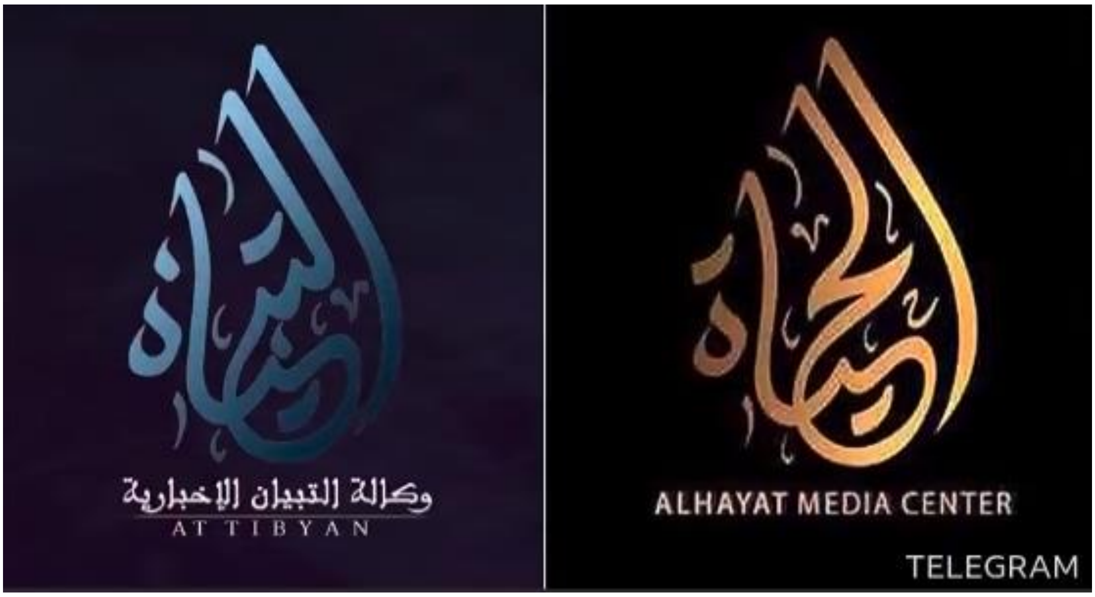


Figure 29: Boko Haram’s Attibyan media entity was designed similarly to IS’s al-Hayat Media Centre.[92]

[90] https://jihadintel.meforum.org/identifier/514/boko-haram-al-urwa-al-wuthqa

[91] https://monitoring.bbc.co.uk/product/c200exu9

[92] https://monitoring.bbc.co.uk/product/c200exu9

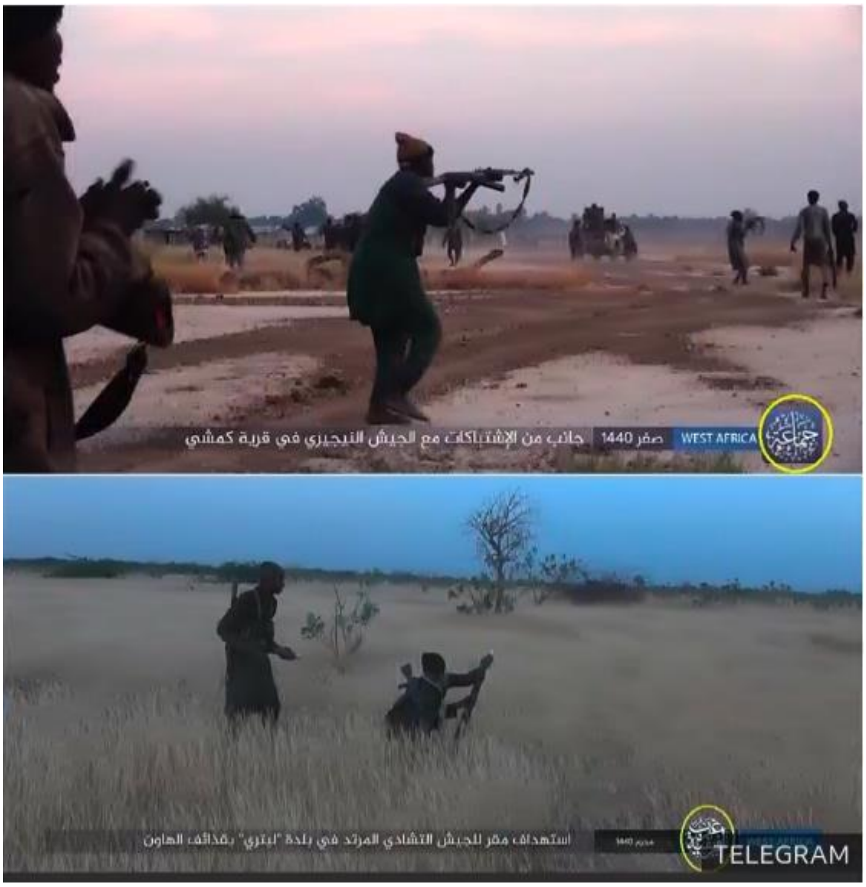

WEST AFRICA
صفر 1440
جانب من الإشتباكات مع الجيش النيجيري في قرية كمشي
محرم 1440
استهداف مقر للجيش التشادي المرتد في بلدة "ليتري" بقذائف الهاون
TELEGRAM

Figure 30: Boko Haram (top) propaganda was styled similarly to Islamic State’s propaganda content (bottom).[93]

[93] https://monitoring.bbc.co.uk/product/c200exu9

Case Study: Al-Shabaab

Background

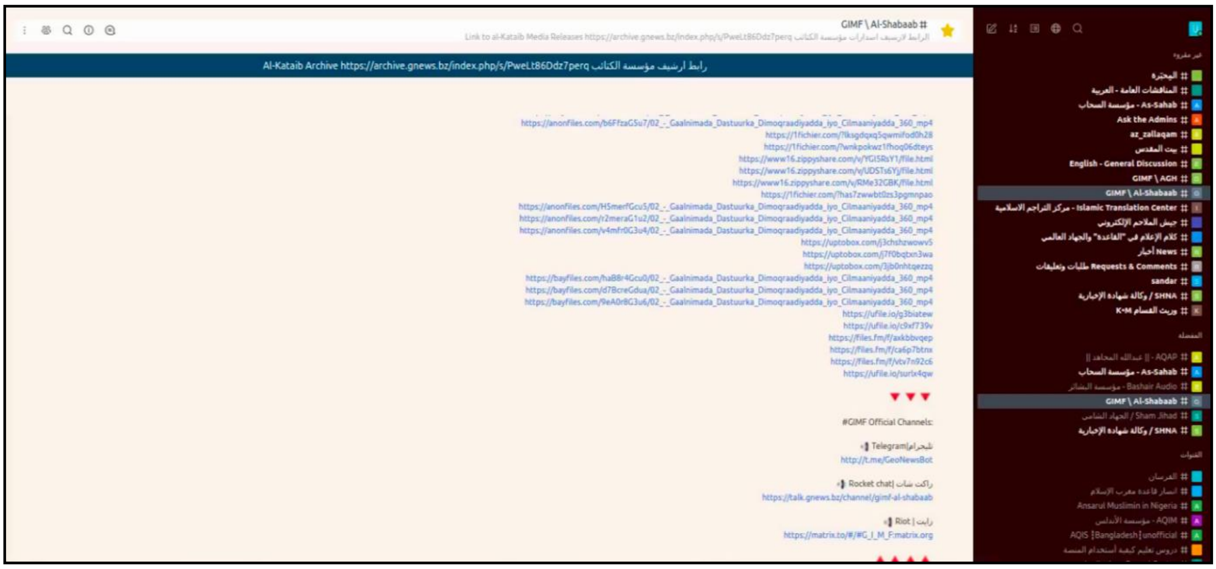
Al-Shabaab (also known as Harakat al-Shabaab al-Mujahideen, حركة الشباب المجاهدين, Xarakada Mujaahidiinta Alshabaab) is a violent Islamist militant organisation based primarily in Somalia, as well as in Kenya and, to a lesser extent, in Ethiopia. The group formed in 2006 after breaking away from the Islamic Courts Union in Somalia, a loose association of Islamic courts between 2000-2007.[94] The group has been designated as a terrorist organisation by the US since 2008, and was officially accepted as an affiliate of al-Qaeda in 2012. The group's most deadly attacks have included a truck bombing in Mogadishu in October 2017, which killed more than 580 people, and assaults on a University and Hotel in Kenya in 2013 and 2019 respectively. Its current leader is Ahmed Umar Abu Ubaidah.

Al-Shabaab maintains a complex and coordinated online presence, spanning several distinct outlets operating across multiple online platforms. The group's messaging operates both locally and internationally, reaching rural Somali communities with limited internet access via traditional radio stations, and communicating more widely via websites and international social media and messaging platforms. Central to the group's dissemination strategy is its primary media arm, al-Kataib Foundation, which produces all its officially branded content, often in conjunction with the al-Qaeda-affiliated Global Islamic Media Front (GIMF).

The group operates a multiplatform approach to disseminating content across the internet, utilising several simultaneous "beacon" channels on messaging apps to share content in long lists of outlinks. Al-Shabaab is a particularly prolific user of outlinks, often uploading new propaganda videos or other multimedia content to tens of platforms simultaneously, sharing hundreds of unique URLs to the content in its beacon platforms or on paste sites. This tactic reflects the group's deliberate shift in their online activities, as both larger and smaller tech platforms make improvements in removing or otherwise moderating their content.

As an example, an al-Shabaab video released in early September 2021 by al-Kataib Foundation was uploaded to at least 111 separate locations on 19 platforms, including versions in three different pixel resolutions. The group posted an aggregated list of URLs to copies of the video in its channel on Geonews, an al-Qaeda-affiliated platform. The use of long URL lists is intended to maximise the time that the content is available for the group's followers; the video remains accessible, as it did in this case, for as long as it takes the slowest platform to take it down.

[94] https://cisac.fsi.stanford.edu/mappingmilitants/profiles/islamic-courts-union



GIMF \ Al-Shabaab ⌗
Link to al-Kataib Media Releases https://archive.gnews.bz/index.php/s/PweLt86Ddz7perq الرابط لارشيف اصدارات مؤسسة الكتائب

Al-Kataib Archive https://archive.gnews.bz/index.php/s/PweLt86Ddz7perq رابط ارشيف مؤسسة الكتائب

https://anonFiles.com/b6FfzaG5u7/02_-_Gaalnimada_Dastuurka_Dimoqraadiyadda_iyo_Cilmaaniyadda_360_mp4
https://1fichier.com/?lksgdqxq5qwmifod0h28
https://1fichier.com/?wnkpokwz1fhoq06dteys
https://www16.zippyshare.com/v/YGI5RsY1/file.html
https://www16.zippyshare.com/v/UDSTs6Yj/file.html
https://www16.zippyshare.com/v/RMe32GBK/file.html
https://1fichier.com/?has7zwwbt0zs3pgmnpao
https://anonfiles.com/H5merfGcu5/02_-_Gaalnimada_Dastuurka_Dimoqraadiyadda_iyo_Cilmaaniyadda_360_mp4
https://anonfiles.com/r2meraG1u2/02_-_Gaalnimada_Dastuurka_Dimoqraadiyadda_iyo_Cilmaaniyadda_360_mp4
https://anonfiles.com/v4mfr0G3u4/02_-_Gaalnimada_Dastuurka_Dimoqraadiyadda_iyo_Cilmaaniyadda_360_mp4
https://uptobox.com/j3chshzwowv5
https://uptobox.com/j7f0bqtxn3wa
https://uptobox.com/3jb0nhtqezzq
https://bayfiles.com/haB8r4Gcu0/02_-_Gaalnimada_Dastuurka_Dimoqraadiyadda_iyo_Cilmaaniyadda_360_mp4
https://bayfiles.com/d7BcreGdua/02_-_Gaalnimada_Dastuurka_Dimoqraadiyadda_iyo_Cilmaaniyadda_360_mp4
https://bayfiles.com/9eA0r8G3u6/02_-_Gaalnimada_Dastuurka_Dimoqraadiyadda_iyo_Cilmaaniyadda_360_mp4
https://ufile.io/g3biatew
https://ufile.io/c9xf739v
https://files.fm/f/axkbbvqep
https://files.fm/f/ca6p7btnx
https://files.fm/f/vtv7n92c6
https://ufile.io/surlx4qw

▼▼▼

#GIMF Official Channels:

Telegram|تليجرام
http://t.me/GeoNewsBot

Rocket chat| راكت شات
https://talk.gnews.bz/channel/gimf-al-shabaab

Riot | رايت
https://matrix.to/#/#G_I_M_F:matrix.org

غير مقروءة
⌗ المحبّرة
⌗ المناقشات العامة - العربية
⌗ As-Sahab - مؤسسة السحاب
⌗ Ask the Admins
⌗ az_zallaqam
⌗ بيت المقدس
⌗ English - General Discussion
⌗ GIMF \ AGH
⌗ GIMF \ Al-Shabaab
⌗ Islamic Translation Center - مركز التراجم الاسلامية
⌗ جيش الملاحم الإلكتروني
⌗ كلام الإعلام في "القاعدة" والجهاد العالمي
⌗ News أخبار
⌗ Requests & Comments طلبات وتعليقات
⌗ sandar
⌗ SHNA / وكالة شهادة الإخبارية
⌗ K•M وريث القسام
المفضلة
⌗ AQAP - || عبدالله المجاهد ||
⌗ As-Sahab - مؤسسة السحاب
⌗ Bashair Audio - مؤسسة البشائر
⌗ GIMF \ Al-Shabaab
⌗ Sham Jihad / الجهاد الشامي
⌗ SHNA / وكالة شهادة الإخبارية
القنوات
⌗ الفرسان
⌗ انصار قاعدة مغرب الإسلام
⌗ Ansarul Muslimin in Nigeria
⌗ AQIM - مؤسسة الأندلس
⌗ AQIS ¦Bangladesh¦unofficial
⌗ دروس تعليم كيفية أستخدام المنصة

Figure 31: An al-Shabaab video release on GeoNews in September 2021, featuring long lists of URLs to copies of the video across multiple platforms.

As the below graph shows, al-Shabaab most frequently targets smaller file sharing platforms for the sharing of its material immediately after release, likely with the use of a file mirroring service such as Mirrorace. File mirroring services allow users to upload single pieces of content to multiple platforms simultaneously, and often include in their services platforms such as those included in the graphic below.

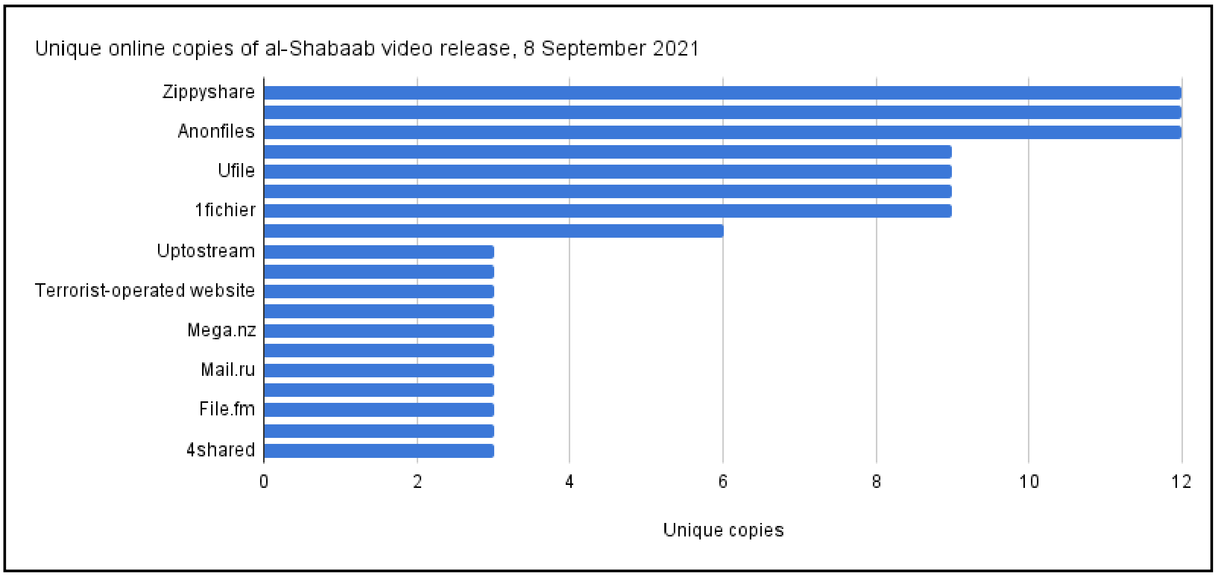
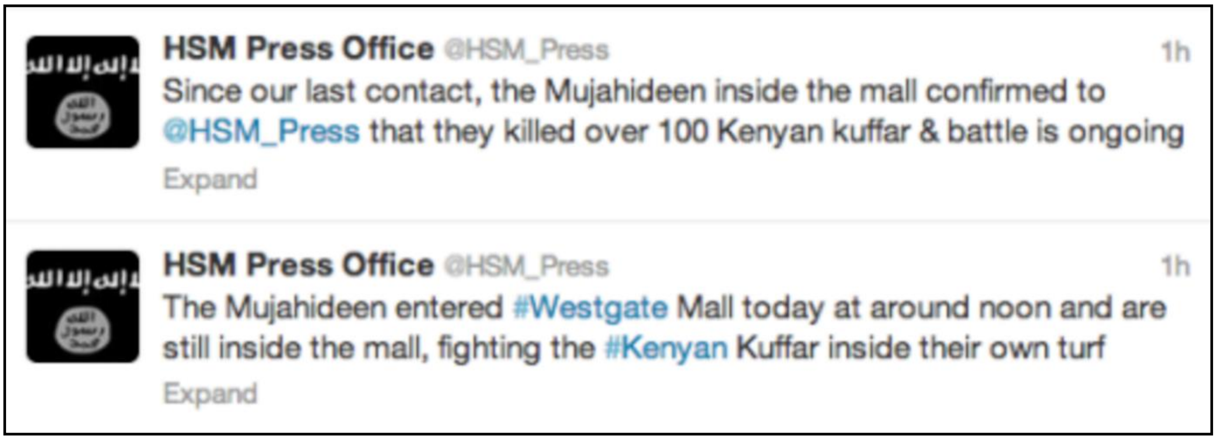


Figure 32: Breakdown of unique copies of the video, September 2021 (source: Terrorist Content Analytics Platform (TCAP)).

Key Propaganda Outlets

HSM Press Office

	The HSM Press Office, short for Harakat al-Shabaab al-Mujahideen Press Office, is an al-Shabaab unit responsible for external communications, particularly official statements. The unit notably ran official accounts on Twitter in the early 2010s, including live-tweeting a deadly attack on a shopping centre in Nairobi, Kenya in 2013. HSM Press Office no longer maintains a presence on Twitter, but its logo appears on all official communications from the group including attack claims and leadership statements released in partnership with al-Kataib Foundation or the Global Islamic Media Front (GIMF).



HSM Press Office @HSM_Press 1h
Since our last contact, the Mujahideen inside the mall confirmed to @HSM_Press that they killed over 100 Kenyan kuffar & battle is ongoing
Expand

HSM Press Office @HSM_Press 1h
The Mujahideen entered #Westgate Mall today at around noon and are still inside the mall, fighting the #Kenyan Kuffar inside their own turf
Expand

Figure 33: Screenshot of tweets posted by HM Press Office during al-Shabaab attack on Westgate mall in Nairobi, Kenya, in 2013.

Al-Kataib Foundation

Key platforms of operation: Rocketchat, Terrorist operated websites (TOWs), Telegram, WhatsApp

Al-Kataib Foundation (Arabic: مؤسسة الكتائب; Somali: Mu'asasada Al Kataa'ib) is al-Shabaab's official producer of propaganda outlet, with its logo featuring on all official al-Shabaab releases including statements, videos and photo reports. Al-Kataib publishes material for a global audience in several languages including Somali, Arabic, English and Swahili, and has frequently featured content from other al-Qaeda affiliates in its material, including Jamaat Nusrat al-Islam wal Muslimeen (JNIM)'s al-Zallaqa Foundation, al-Qaeda in the Arabian Peninsula (AQAP)'s al-Malahim Foundation and al-Qaeda central's as-Sahab Foundation. These al-Qaeda-affiliated media organisations have also periodically released coordinated statements alongside al-Kataib Foundation. Several of them participated in a messaging campaign beginning in 2019 called "Jerusalem will never be judaised", for example, in which coordinated statements were released as part of messaging around terrorist attacks.

Al-Kataib Foundation frequently produces a large volume of multimedia propaganda content including videos showcasing al-Shabaab attacks, productions covering the group's political activities in Somalia, feature-length documentaries and photosets. Official al-Kataib releases are typically published on tens or even hundreds of platforms simultaneously, with long aggregated lists of URLs to copies of the content shared in official channels run either by al-Shabaab or the al-Qaeda-linked propaganda disseminator the Global Islamic Media Front (GIMF).

The primary source of al-Kataib-produced videos is on the al-Qaeda-aligned "GeoNews" server, located at talk.gnews.bz. The platform is developed using open-source Rocketchat software, functioning essentially as a chat application with a range of channels devoted to different al-Qaeda affiliates, discussion topics, and languages. Al-Kataib videos are also often released via beacon channels on Element Messenger and Telegram. An online up-to-date archive of al-Kataib content is also accessible via a password-protected cloud platform called “Kataibdrive”; the platform is built using open-source NextCloud software.

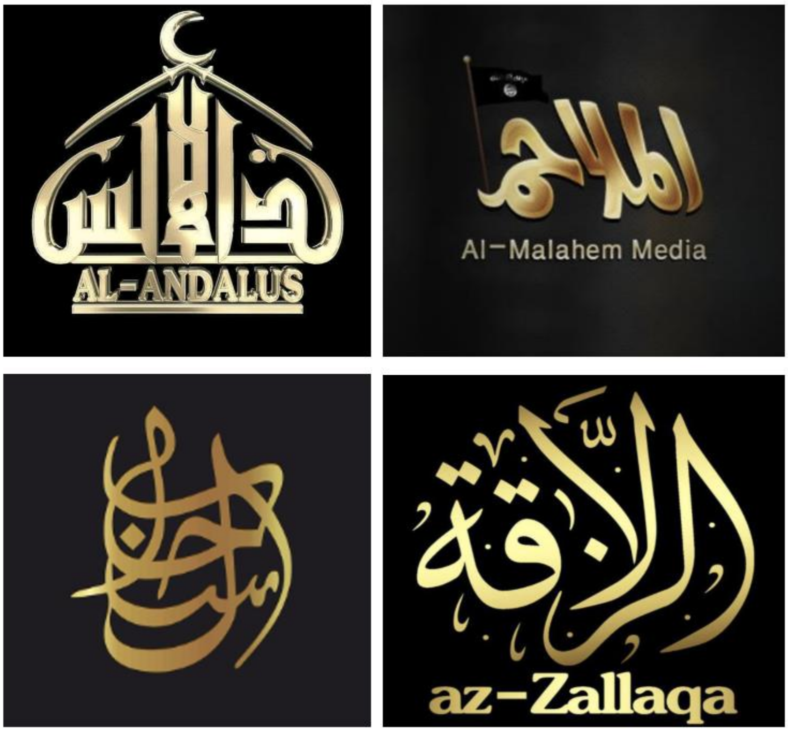


Figure 34: Official media outlets of al-Qaeda and its official affiliates: Bottom left, As-Sahab Foundation, the outlet of al-Qaeda central; Bottom right, az-Zallaqa Foundation of Jama’at Nusrat al-Islam wal Muslimeen (JNIM); Top left al-Andalus Foundation of al-Qaeda in the Islamic Maghreb (AQIM); Top right, al-Malahem Foundation of al-Qaeda in the Arabian Peninsula (AQAP).

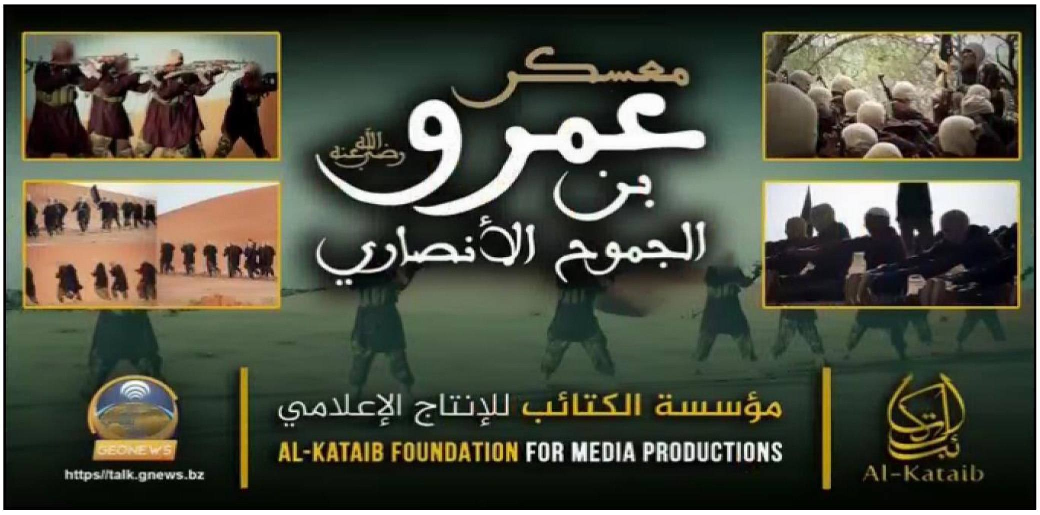


Figure 35: A promotional banner for an al-Shabaab propaganda video produced by al-Kataib, complete with branding on bottom right; sourced from Rocketchat.

Global Islamic Media Front

Key platforms of operation: Rocketchat, Element Messenger, Telegram Messenger, WhatsApp

The Global Islamic Media Front (GIMF) (Arabic: al-Jabhat al-I'lamiyya al-Islamiyya al-'Alamiyya, الجبهة الإسلامية الإسلامية) is an al-Qaeda-aligned propaganda disseminator that produces, releases and promotes online content on behalf of al-Qaeda and several of its official affiliates, including al-Shabaab. GIMF productions are either original, translated or repackaged propaganda, including claims of responsibility for attacks, leadership messages, photosets, or videos. Noteworthy releases such as videos or leadership statements are typically converted into several pixel resolutions and with different language subtitles before being disseminated via long lists of outlinks to content on several file sharing and cloud platforms.

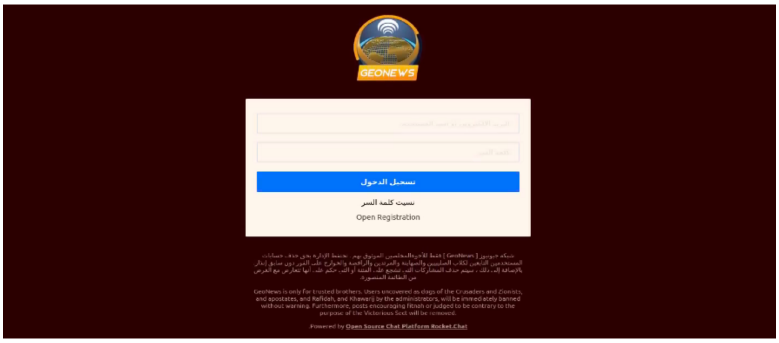
The GIMF primarily publishes al-Shabaab content via two channels on GeoNews, a password-protected pro-al-Qaeda Rocketchat server hosted on its own domain, https://talk.gnews.bz, but also maintains a presence on several other online platforms including Telegram, WhatsApp, Chirpwire and Element Messenger.

GeoNews

Key platforms of operation: Rocketchat, TOWs

GeoNews is a pro-al-Qaeda platform that has been hosted on its own domain since late 2019. The platform is built on open-source Rocketchat code and requires a password to access. It is comprised of multiple messaging groups and channels, each devoted to official and unofficial al-Qaeda groups and topics, with content and discussions in multiple languages. The platform's creation roughly coincided with a Europol-led operation to deplatform violent Islamist actors, including those affiliated with al-Shabaab, from Telegram in November 2019. The operation had the unintended consequence of dispersing terrorist actors across the web and onto several niche alternative apps including Element, Riot and Threema. Such actors still maintain a presence on these apps, but GeoNews has become the most stable, centrally located hub for propaganda dissemination and recruitment.

Al-Shabaab content on GeoNews is primarily disseminated via a joint channel with the GIMF. All content produced by al-Kataib Foundation is posted here. Another GIMF-run channel shares news, including attack claims, relating to al-Shabaab.



شبكة جيونيوز [GeoNews] فقط للأخوةالمخلصين الموثوق بهم . تحتفظ الإدارة بحق حذف حسابات المستخدمين التابعين لكلاب الصليبيين والصهاينة والمرتدين والرافضة والخوارج على الفور دون سابق إنذار. بالإضافة إلى ذلك ، سيتم حذف المشاركات التي تشجع على الفتنة أو التي حكم على أنها تتعارض مع الغرض من الطائفة المنصورة.

GeoNews is only for trusted brothers. Users uncovered as dogs of the Crusaders and Zionists, and apostates, and Rafidah, and Khawarij by the administrators, will be immediately banned without warning. Furthermore, posts encouraging fitnah or judged to be contrary to the purpose of the Victorious Sect will be removed.

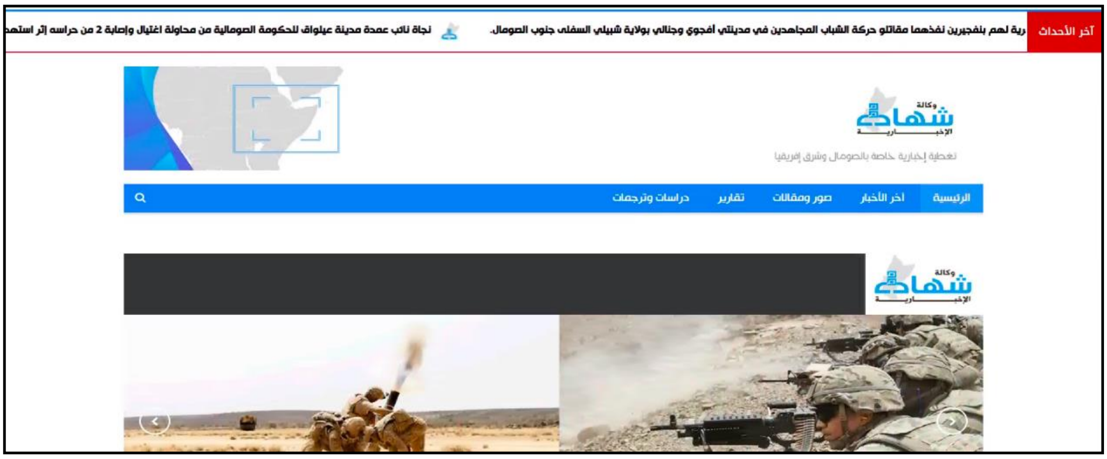
.Powered by Open Source Chat Platform Rocket.Chat

Figure 36: A login page on the home screen of GeoNews, August 2021.

Shahada News Agency

Key platforms of operation: Telegram, TOW, GeoNews

Shahada News Agency (Arabic: وكالة شهادة الإخبارية) is an al-Shabaab-run news outlet that primarily produces Arabic-language news relating to the group’s operations and general current affairs in the wider East Africa region. The outlet attempts to present itself in its reporting as an independent organisation, but it often acts as the first point of release for al-Shabaab statements and attack claims. Its other reporting on the region is generally framed in support of al-Shabaab and includes coverage of the group’s allegedly positive work in Somalia, such as in public order, or the building and maintenance of infrastructure. The organisation’s most prominent online channels for content dissemination are its website, shahadanews.com, and a bot on Telegram at @Shahaada_bot. Its messaging around al-Shabaab attacks is also republished by the GIMF in an aggregate channel on GeoNews.



آخر الأحداث

رية لهم بلفجيرين نفذهما مقاتلو حركة الشباب المجاهدين في مدينتي أفجوي وجنالي بولاية شبيلي السفلى جنوب الصومال.

نجاة نائب عمدة مدينة عيلواق للحكومة الصومالية من محاولة اغتيال وإصابة 2 من حراسه إثر استهد

تغطية إخبارية خاصة بالصومال وشرق إفريقيا

Figure 37: Shahada News Agency’s website, August 2021.

Somalimemo

Key platforms of operation: TOWs, Facebook, Twitter

Somalimemo is a Mogadishu-based pseudo-news outlet that operates in support of al-Shabaab. Like Shahada News Agency, Somalimemo purports to be an independent, non-partisan news organisation in its reporting, but its coverage of Somali news and al-Shabaab operations is heavily biased in support of the group, almost always reporting its ideological message and perspective, including inflated casualty figures from al-Shabaab attacks and articles that promote anti-government sentiment.

Given Somalimemo’s sophisticated attempts at distancing itself from al-Shabaab in its public messaging, it often evades moderation by tech platforms. At the time of writing its primary channels for content dissemination were on its website, currently located at somalimemo.info, and associated accounts on Facebook and Twitter. Tech Against Terrorism facilitated the suspension in early 2021 of two of the group’s earlier domains, Somalimemo.net and Somalimemo24.com.

Bogga Calamada

Main platforms of operation: TOWs, Telegram, Google Podcasts

Bogga Calamada is one of al-Shabaab’s more peripheral supporter-run “news” channels. It poses as a legitimate and independent news organisation, reporting on a range of issues relating to Somalia, East Africa and internationally, but its coverage is overwhelmingly weighted towards a favourable view of al-Shabaab’s operations and ideological project.

Calamada’s primary point for content dissemination is via its website, Calamada.com, as well as a channel on Telegram. The site is regularly updated with articles and other multimedia content, including propaganda originally published by al-Shabaab’s official media outlets.

Radio al-Andalus

Key platforms of operation: Local radio, TOWs, Telegram

Radio al-Andalus (Idaacadda al-Andalus) is a radio station based in Somalia and run by al-Shabaab. Al-Shabaab’s control and operation of radio stations within Somalia has over the past several years been as a result of the capture of multiple relays within including HornAfrik and Holy Koran Radio. Radio al-Andalus broadcasts propaganda both via analogue radio and online relating to the group, including reports of its operations, messaging, official statements, and Islamic music dedicated to the group’s extremist interpretation of Islam. Through analogue radio broadcasts, al-Shabaab reaches populations within Somalia and the Horn of Africa that may not have access to a stable internet connection.

Radio al-Andalus is broadcast online via its dedicated website, somalimp3.net, and is promoted by al-Shabaab-affiliated entities Bogga Calamada and Somalimemo. Whois information listed with regards to both Radio al-Andalus, Somalimemo and Radio al-Furqaan, another official al-Shabaab radio station, indicates links to a company based in Mogadishu called Waasuge Media Group.

Radio al-Furqaan

Key platforms of operation: TOWs, Telegram Messenger

Radio al-Furqaan (Somali: Idaacadda Al-furqaan) is another official al-Shabaab-run media outlet and radio station based in Somalia. Like Radio al-Andalus the station broadcasts propaganda about the group’s activities and general news in Somalia, Eastern Africa and globally, often with a focus on al-Shabaab’s alleged social development programs.

Radio al-Furqaan operates its own website and a channel on Telegram Messenger, where it had almost 2,200 subscribers at the time of writing. It posts between 3-10 times per day, on average.

Thabat News Agency

Key platforms of operation: TOWs, Telegram Messenger

وكالة إخبارية
ثبات

Thabat News Agency (وكالة ثبات الإخبارية) is a pro-al-Qaeda propaganda outlet and news aggregator that reports on activities of al-Qaeda and its affiliates, including al-Shabaab. The group publishes a regular weekly “newsletter” covering global operations, and its online channels circulate content relating to affiliates through photos, attack claims, and infographics summarizing alleged attacks over time.

Thabat Agency runs a bot on Telegram and transitory websites often utilizing free web design software such as site123. In July 2021 the group promoted its own app in .apk format, on which it said supporters could access its content in a curated and centralized fashion via Android.

Dossier B: Legal Responses to Terrorist Use of the Internet

Executive Summary

Legal responses to terrorist use of the internet in African Commonwealth countries

Few of the African Commonwealth member countries have laws specifically targeting terrorist content and terrorist use of the internet.

Sierra Leone, Kenya and Nigeria are exceptions, as they prohibit terrorist use of the internet or the dissemination of terrorist content in statutes concerned primarily with cybercrime. Mauritius' Prevention of Terrorism Act[95] also includes provisions on blocking individual access to the internet to prevent a terrorism act.

Certain existing provisions to regulate illegal online content include limitations on content harmful to national security and public order, yet very few target the diffusion of terrorist content specifically.

While such laws could, if amended, provide a more rigorous legal framework to counter terrorist activity online, any amendments should be clearly drafted, detail unambiguously what constitutes terrorist activity, and include proper safeguards for the rule of law and human rights. This is discussed under the heading 'recommendations' below.

Kenya, Lesotho, Mauritius, Sierra Leone, Uganda, and Tanzania have all begun to discuss to introduce laws to regulate online content and the use of social media in the last 3 years.[96] However, much of this regulation concerns illegal content in general, or misinformation and disinformation, rather than terrorist use of the internet. When proposing legislation to limit online content, some countries in the region such as Rwanda[97] and Uganda[98] have both cited laws proposed or passed in Europe to argue for regulation of online spaces.

The lack of regulation concerned exclusively with terrorist content and use of the internet, as well as the emergence in the last few years of new regulation of online spaces generally, is not specific to the region: such developments conform to a global trend of increased regulation of the online space analysed by Tech Against Terrorism in the Online Regulation Series.[99]

[95] https://www.ilo.org/dyn/natlex/docs/ELECTRONIC/104041/126736/F1839294808/MUS104041.pdf

[96] For Kenya see: proposed Kenya Information and Communication (Amendment) Bill, 2019, http://www.parliament.go.ke/sites/default/files/2019-06/Kenya%20Information%20and%20Communications%20%28Amendment%29%20Bill%2C%202019.pdf;
For Lesotho see: https://www.voanews.com/africa/tiny-african-nation-lesotho-proposes-social-media-limits;
For Mauritius see: proposed Amendments to the ICT Act for Regulating the Use and Addressing the Abuse and Misuse of Social Media, https://www.icta.mu/docs/2021/Social_Media_Public_Consultation.pdf;
For Sierra Leone see: Cyber Crime Act 2020 (passed in June 2021), http://www.sierra-leone.org/Laws/2020-Cybercrime%20Act.pdf;
For Uganda see: https://uccinfo.blog/2019/07/19/regulation-and-responsible-use-of-online-media/;
For Tanzania see: Electronic and Postal Communications (Online Content) Regulations, 2020, https://thrdc.or.tz/wp-content/uploads/2019/09/ONLINE-CONTENT-REGULATIONS.pdf

[97] https://www.newtimes.co.rw/news/govt-moves-regulate-social-media-content-amid-misinformation

[98] https://uccinfo.blog/2019/07/19/regulation-and-responsible-use-of-online-media/

[99] In the 17 jurisdictions analysed by Tech Against Terrorism for the Online Regulation Series Handbook, more than 25 24 statutes, legislative amendments and executive measures have been promulgated or proposed since 2017 to regulate harmful online content or counter terrorist content online. Read the full Handbook here: https://www.techagainstterrorism.org/2021/07/16/the-online-regulation-series-the-handbook/

Global trends in online regulation and legal responses to terrorist use of the internet:

In the past four years, global policymakers have increasingly sought to regulate online content. Tech Against Terrorism has identified three separate regulatory aims used by governments to justify regulating the online space:

- Countering terrorist and violent extremist content, or more broadly "harmful" content.
- Countering the spread of misinformation and disinformation.
- Adapting to the risks of today's digital world.

The first category of stated regulatory aims, on countering terrorist and violent extremist content, often focuses on countering the dissemination of terrorist content by mandating tech platforms to prevent the dissemination of such content and rapidly remove it.

Case Study: Germany

Germany was one of the first countries to require tech platforms to remove terrorist content within a short timeline (24-hour) with the Network Enforcement Act passed in 2017. Since then, the European Union has adopted Regulation 2021/784 on Addressing the Dissemination of Terrorist Content Online in 2021, mandating a one-hour removal deadline; and the initial "cyberhate" law in France also included a one-hour removal deadline for terrorist and child sexual abuse content. Commentators anticipate that forthcoming online regulations in Canada will also include a 24-hour removal deadline for illegal content.

In its analysis of regulations aimed at countering the dissemination of terrorist content, and in certain jurisdictions of other "harmful", Tech Against Terrorism identified the following regulatory trends:

- Short removal deadlines requiring platforms to remove content within a short timeframe, often ranging from 1 to 48 hours, following a notice from the relevant authorities. Such measures have been introduced in several European countries and at the EU level, as well as in India, Pakistan, Australia, Turkey, and Indonesia.
- Outsourcing legal adjudication to tech platforms, with companies required by law to assess whether content is illegal following a report from authorities, or in certain instances from users. This is predominantly a trend in Europe but is also in Australia's Online Safety Act.
- Incentivising increased reliance on automated content moderation. This often stems from short removal deadlines, however, some countries, such as Pakistan, explicitly require platforms to deploy proactive moderation tools or to prevent the re-upload of previously removed content and livestream of terrorist and violent extremist content.
- Mandating transparency and accountability measures from platforms. This has mainly been brought forward in Europe (but also in India and Australia) via requirements for platforms to produce transparency reports on their compliance with new regulations and removal of illegal content.
- Disproportionate legal requirements for smaller platforms by applying regulations indiscriminately to platforms of all sizes, with the underlying expectation that smaller platforms should comply with the same capacity as larger ones to stringent requirements.

Note: The analysis of global online regulation and legal responses to terrorist use of the internet included in this report, are based on Tech Against Terrorism's Online Regulation Series.[100]

This analysis has been tailored to provide a high-level summary of best practices and key concerns with emerging trends in online regulation, as well as to provide an overview of legal responses to terrorist use of the internet in Commonwealth countries in Sub-Saharan Africa. For a more detailed survey of the state and future of online regulation in 17 jurisdictions, including regulatory key

[100] https://www.techagainstterrorism.org/2020/12/22/the-online-regulation-series-summary/

trends and Tech Against Terrorism's recommendations for governments, please see The Online Regulation Series Handbook.[101]

Tech Against Terrorism's Recommendations for Governments

Having analysed existing and forthcoming regulations intended to counter harmful and terrorist content online, Tech Against Terrorism calls on governments to uphold due process and the rule of law by making the following commitments

1 Safeguard the rule of law

- Ensure that definitions of key terms, such as terrorist content, are clear, practical, and have a basis in existing legal frameworks.
- Avoid introducing regulation that depends on subjective interpretations of harm, as these can be difficult for tech companies to implement at scale without negatively impacting freedom of expression.
- Refrain from criminalising online what is legal offline. Governments should provide a clear legal basis for requesting platforms to remove content, including via existing counter-terrorism laws and terrorist designation lists which are critical to informing platforms' counter-terrorism approach.[102]
- There should be a clear legal basis for removing online content via existing counter-terrorism laws and terrorism designation lists.
- Protect freedom of expression in line with international human rights standards, by reserving adjudication on its lawful limits to an independent judiciary rather than an administrative body.
- Provide legal certainty to tech platforms by clarifying how regulatory compliance will be assessed, and by providing guidance on the specific steps that companies should take to comply with legal requirements

2 Honour commitments to due process when implementing online regulations:

- Provide transparent accounts of the steps taken by regulatory bodies in the exercise of their authority. This allows for public assessment of the extent to which such bodies are:
    - fully aware of risks to human rights and freedom of expression associated with content moderation measures and information requests.
    - consistent in their application of the law and free of political bias in making removal orders.
    - consistent and accurate in issuing penalties to commercial providers.
    - free of incentive to be overzealous in moderating content.
    - accountable for their operational assessments and judgements
- Clarify safeguards and redress mechanisms for users by stating:
    - What safeguards are in place to prevent the removal of legal content.
    - How erroneous removal can be remedied, particularly in cases where removal has been requested by a country's judicial or governmental authority.

3 Produce transparency reports on governmental engagement with tech companies for counter-terrorism purposes

- Transparency reports in line with the Tech Against Terrorism Guidelines on Transparency Reporting on Counter-terrorism Efforts[103]

[101] https://www.techagainstterrorism.org/2021/07/16/the-online-regulation-series-the-handbook/
[102] https://www.techagainstterrorism.org/2020/05/26/the-designation-of-the-russian-imperial-movement-by-the-us-state-department-why-it-matters-for-tech-companies/
[103] https://transparency.techagainstterrorism.org/

4 Consider the capacity and resources of smaller platforms and uphold the principles of proportionality in regulation and equality before the law.[104]

- Ensure obligations for tech companies are proportionate to size and capacity and promote competition and innovation by limiting financial penalties for smaller or micro-platforms.
- Increase support for the tech sector, for smaller platforms in particular, in countering terrorist and violent extremist use of the internet - for example, through public-private partnership endeavours, and digital literacy programmes.

5 Exclude measures which pose a risk to freedom of expression, diversity of content, and technical innovation whose efficacy for tackling terrorist use of the internet is unproven:

- Any provision which does not allow sufficient time for platforms to adequately assess the legality of content, nor provide the necessary practical support for platforms to make assessments correctly.
    - Included in this is the requirement for platforms to remove terrorist and other harmful content by a set deadline, which does not account for platforms' capacities and encourage the overzealous removal of content.
- Any attribution of liability for user-generated content to tech companies or their employees, which penalises those trying to counter terrorist content rather than those disseminating that content.

Tech Against Terrorism calls on governments to take a holistic approach to countering terrorism and violent extremism. In addition to regulating terrorist and harmful content, governments should ensure that regulatory frameworks address the root causes of radicalisation and hold accountable those individuals that engage in terrorist and violent extremism activities, in full compliance with international human rights standards.

Emerging Online Regulation – Good Practice

Acknowledge smaller platforms' limited resources and capacity

Tech Against Terrorism's analysis of the state of global online regulation shows that most regulations proposed or passed in the last four years apply indiscriminately to platforms of all sizes, without consideration for the difference in resources and capacity to comply with stringent legal requirements.

However, some online regulatory approaches include provisions only relevant to larger platforms. Whilst differentiated requirements depending on a platform's size are often limited to provisions related to transparency and accountability, they nevertheless signal an acknowledgement of the fact that smaller platforms should not be expected to satisfy the same demanding requirements as larger ones.

Policymakers should further acknowledge the differences in platforms' resources and capacities and draft legislation accordingly, for instance by allowing more time to smaller platforms to adapt their processes and systems to new legislation or by providing them with the support needed to comply.

Tech Against Terrorism also recommends that policymakers clarify in the regulatory frameworks the categorisation of platform size and consider not only the userbase but also platforms' resources (financial, human and technical) in the categorisation process. This would ensure that platforms lacking resources are not mistakenly categorised.

[104] Concerns regarding disparities in resources and how these impact a platform's capacity to comply with legal requirements were also raised by the French Constitutional Council in its censure of the so-called "Cyberhate" law. The Council stressed that some of the provisions in the original draft were impossible to satisfy and breached the principle of 'l'égalié devant les charges publiques' which demands that legal and administrative requirements should not place heavy or discriminatory burdens on those having to comply. With this ruling, the Council recognised that platforms' resources can significantly impede their compliance with legal requirements, and that requirements with high resource demands should not be included in online regulation.

Example: Online regulations in India and Turkey, include specific compliance requirements for large platforms. However, the definition of what constitutes a "large" platform is not always clear in such laws; further clarification is therefore sought from those authorities charged with overseeing their implementation. The Bill on Separatism in France also requires platforms whose user-base exceeds a certain size to comply with specific requirements on countering the spread of "illegal and hateful content", including by review of their algorithms

The Platform Accountability and Consumer Transparency Act (PACT Act),[105] one of the many proposals to reform section 230 [106] in the US,[107] further acknowledges the importance of varying expectations depending on a platform's size. In particular, the Act outlines requirements which are scalable in accordance with a platform's revenue.

Supporting transparency and accountability

Commendably, the majority of online regulations introduced in 2019-2021 include provisions that seek to increase transparency and accountability from the tech sector – whether by mandating detailed Terms of Services[108] or increased transparency reporting.[109]

Detailed Terms of Services and Community Guidelines, which clearly explains what is allowed or not on the platform, and how violating content or behaviour will be actioned, are crucial to ensure accountability. Clear and detailed Terms of Services publicly inform users of the ground rules of content moderation and act as the reference document for users to be able to understand why content was actioned or how to contest a removal decision if they think their content was removed in error. In the last four years, provisions mandating platforms to have clear and detailed Terms of Services have become increasingly common in regulation aimed at countering illegal and harmful content.[110]

Example: EU Regulation 2021/784[111] states that platforms should clearly prohibit terrorist content in their community guidelines, whereas the EU Digital Service Act[112] and the UK Guidance for Video Sharing Platforms[113] outline what platforms should include in their content

[105] The PACT Act was originally introduced in the 2019 – 2020 Congressional Session and reintroduced with significant amendments in 2021. This proposal is bipartisan, supported by Senators Brian Schatz (Democrat) and John Thune (Republican).

[106] https://slate.com/technology/2021/03/section-230-reform-legislative-tracker.html

[107] Section 230 of the Communication Decency Act of 1996, establishes intermediary liability protections related to user-generated content in the US, meaning that tech companies are not seen as liable for content posted by their users.

[108] Terms of Service (ToS) are rules with which one must observe and abide if they wish to use the service. It is a legally binding agreement, required by all platforms that store personal data for a user. The relevant documentation contained in a platform's ToS varies in response to the service that a platform offers. Social media platforms will pair their Terms of Service with Community Guidelines. Financial Technology ('Fintech') will contain a Privacy Policy. Storage platforms, such as iCloud, might have an Acceptable Use Policy.
For most platforms offering user-to-user services (including communications, content hosting and sharing), the Community Guidelines (also known as User Guidelines) are the most important document of the ToS as they outline what is acceptable or not on a platform. In general, Community Guidelines are to be understood by users as a set of foundational principles aiming to balance self-expression with safety, to protect both users and platform.

[109] Transparency is vital to ensure that the tech industry and governments are accountable to the public. Transparency reporting provides insight on to what extent fundamental freedoms such as freedom of expression and the right to privacy are respected across the internet when countering terrorist use of the internet. It can also encourage and recognise meaningful action from tech companies in tackling terrorist use of the internet and provide crucial insight on this threat. You can find Tech Against Terrorism's Guidelines on Transparency Reporting on Online Counter-terrorism Efforts (for tech platforms and governments) here: https://transparency.techagainstterrorism.org/

[110] https://www.techagainstterrorism.org/2021/07/16/the-online-regulation-series-the-handbook/

[111] https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52018PC0640

[112] https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

[113] https://www.ofcom.org.uk/tv-radio-and-on-demand/information-for-industry/vsp-regulation

standards. The 2020 Rules in Pakistan[114] and the 2021 Guidelines[115] in India both go a step further and require platforms to add to their content standards the list of content prohibited by law.

Transparency is also vital to ensure accountability towards the public and internet users, and transparency reporting on counter-terrorism efforts provides insight into the extent to which human rights and fundamental freedoms are respected across the internet in the fight against terrorist use of online platforms.

Tech Against Terrorism thus welcomes government calls for increased transparency and accountability from the tech sector in countering terrorist use of the internet. However, Tech Against Terrorism cautions against mandating transparency reporting to a uniform standard across platforms, as this would disregard the diversity of services offered and the differences in resources and capacity. Tech Against Terrorism recommends that governments support the Guidelines for Transparency Reporting on Online Counter Terrorism Efforts,[116] which focus on a small number of core metrics to facilitate the evaluation of performance over time, and which fully comprehend the importance of platform diversity.

Beyond transparency reporting for smaller platforms, Tech Against Terrorism also recommends that governments publish transparency reports on their online counter-terrorism efforts and collaboration with tech platforms to counter terrorist content. Tech Against Terrorism recommends that governments report on the legal basis of online counter-terrorism activities and on their commitment to international law.

Example: On transparency, the draft Online Safety Bill[117] in the UK demands that platforms publish transparency reports on their compliance with the Bill. EU Regulation 2021/784[118] also requires tech companies to publish transparency reports on their efforts to comply with the regulation, and outlines metrics for transparency reporting by governments and competent authorities. France’s “cyberhate” law also calls for increased transparency from both the tech and government sectors and requires the country’s telecoms authority to publish an annual report on the enforcement of the law.[119]

Risk assessment

Understanding the threat is a crucial first step to effectively counter terrorist use of the internet and the diffusion of terrorist content online. For tech platforms of all sizes and services, this means properly comprehending the threat they face, and the strategies employed by terrorist actors to exploit their services and evade moderation. Despite the obvious need for a proper understanding of the threat, Tech Against Terrorism notes that most tech platforms, particularly those with fewer resources, are unable to achieve this.

Tech Against Terrorism thus welcomes regulatory provisions highlighting the importance of conducting risk assessments which can be of critical assistance to tech companies. Risk assessments permit consideration of how platforms can be exploited. and of what potentially adverse shifts in usage they should remain vigilant. However, Tech Against Terrorism recommends that governments provide the necessary support for all platforms to be able to conduct the required risk assessments. Tech Against Terrorism particularly recommends that governments support public-private partnerships and knowledge-sharing endeavours which help tech companies to understand the threat, such as the Knowledge Sharing Platform developed by Tech Against Terrorism.[120]

[114] https://moitt.gov.pk/SiteImage/Misc/files/CP%20(Against%20Online%20Harm)%20Rules%2C%202020.pdf
[115] https://www.medianama.com/wp-content/uploads/2021/02/Intermediary-Guidelines-2021.pdf
[116] https://transparency.techagainstterrorism.org/
[117] https://www.gov.uk/government/publications/draft-online-safety-bill
[118] https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52018PC0640
[119] https://www.vie-publique.fr/loi/268070-loi-avia-lutte-contre-les-contenus-haineux-sur-internet
[120] https://ksp.techagainstterrorism.org/

Example: The UK draft Online Safety Bill[121] includes a provision mandating platforms to conduct risk assessments regarding the presence of illegal content on their services. The provision stipulates that a new risk assessment must be conducted every time a platform makes a change to policies or operational practices capable of affecting the presence of illegal content on its services. These requirements differ somewhat between providers of user-to-user services and search, and the draft bill lays out criteria to be considered by these different providers when conducting their risk assessments. A similar requirement to conduct risks assessments is already included in the UK Interim Regime for Regulating Video-Sharing Platforms[122] – in effect since November 2020.

Prevent aspects and calls for greater cooperation

Terrorist use of the internet is an ecosystemic problem which cannot be solved by focusing on the exploitation of a single platform. The tech sector, as well as public-private collaboration more generally is thus crucial to efficiently tackling all aspects of terrorist use of the internet.

Online regulation and government-led initiatives to counter terrorist use of the internet should further acknowledge the importance of providing routes to increased cooperation between tech platforms, but also between tech platforms and governments and civil society. Increased cooperation, in particular when government-led, could yield solutions beyond online regulation to tackle counter terrorism and radicalisation both at the roots and as a whole, not just as it occurs online.

Example: The final version of the so-called “cyberhate” law in France[123] mostly retains the “preventative” clauses of the draft approved by the Parliament in May 2021 calling for greater cooperation and increased transparency. The law includes mechanisms for cooperation and information sharing between platforms, though further information on what these tools might be is required.[124]

Emerging Online Regulation – Key Trends and Concerns

Based on Tech Against Terrorism’s analysis of emerging regulation of online activity globally, Tech Against Terrorism has identified some areas of concern.

Lack of consideration for smaller tech companies

Most regulation introduced over the past four years and analysed by Tech Against Terrorism does not adequately account for smaller platforms, especially in considering such platforms’ lack of capacity to comply with regulation. This is problematic when it concerns terrorist content since most terrorist groups exploit smaller platforms for the practical constraints on their capacity to be proactive. To prevent terrorist use of the internet through regulation, lawmakers need to consider capacity constraints.

Concerns for effectiveness

Smaller tech companies face the greatest threat of terrorist exploitation [125] due to terrorist actors’ ability and inclination to exploit their lack of resources or less sophisticated content moderation systems. By drafting legislation with larger platforms in mind and sanctioning smaller platforms

[121] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safety_Bill_Bookmarked.pdf

[122] https://www.ofcom.org.uk/__data/assets/pdf_file/0021/205167/regulating-vsp-guide.pdf

[123] The final version of the “cyberhate” law follows the censuring of its most stringent provisions by the French Constitutional Council , on the ground that they present risks for freedom of expression and lacked consideration for the capacity of all platforms to comply with law.
See: Tech Against Terrorism (2020), Online Regulation Series – France.

[124] For an in-depth analysis of the French “cyberhate” law, please see: https://bit.ly/2VKTN1L.

[125] https://gifct.org/wp-content/uploads/2021/07/GIFCT-TAWG-2021.pdf

instead of offering them the support they needed to counter the threat, the makers public policy will have a counter-productive impact on the problem.

Terrorist actors are sophisticated in their use of technology and have shown resilience in exploiting online platforms, whether to disseminate content, recruit, raise funds, or organise. For instance, terrorists often rely on a multiplatform approach to content dissemination, using file mirroring services to upload content simultaneously across multiple platforms. However, the complex and fast-changing strategies deployed by terrorists to exploit online platforms are all too often absent from the debate and legislative drafting of policy. In practice, this means that the focus of regulation is often misplaced and fails to rise to the challenge posed by terrorist use of the internet – as is the case with policymakers' fixation with 'algorithmic amplification'.[126]

When policymakers consider the removal of safe harbour protection for tech platforms to hold either providers or their employees liable in law for user-generated content, they misplace the burden of responsibility for terrorist use of the internet. Such exposure to liability penalises those acting to counter terrorist content rather than the individuals responsible for sharing it and thus ignores the root causes of terrorist propaganda and use of the internet.

Concerns for human rights and fundamental freedoms

One emerging trend in online regulation globally is the outsourcing of adjudicative functions. from courts and publicly accountable institutions to private and largely unaccountable tech companies, by including provisions requiring platforms to assess whether a content is illegal and to act accordingly.[127] International human rights standards require that the acceptability of limits to freedom of expression should be decided by independent judicial bodies. However, by mandating tech companies to remove content at scale, many online regulations meant to counter online harms instead shift the responsibility of deciding what is harmful and/or illegal content to private entities.[128] Governments outsourcing adjudication of illegality to private tech companies, when this should be the duty of independent judiciaries, risks undermining due process and the rule of law.

There are also clear concerns for freedom of expression when platforms are compelled to remove content within a short timeframe and at risk of penalties and further liability. This artificial choice between rapid content removal or hefty fines means that platforms will lack time to properly adjudicate on the legality or harmfulness of content and are likely to err on the side of over-removal to avoid financial risk.

Online regulation is often impractically broad in the definition of harmful content and circular in the explanation of terrorist content, rarely indicating to tech platforms how to operationalise these definitions when complying with legal provisions. This presents serious risks for freedom of expression, as regulation could be used to pressure tech companies to remove legal or non-violent speech. With such vague definitions of "legal but harmful" content, countries are introducing mechanisms that risk undermining the rule of law by criminalising speech online that is legal offline. Tech Against Terrorism cautions against vague and circular definitions of terrorist or harmful content in laws, and against governments demanding that platforms remove content that is not clearly prohibited by law. Tech Against Terrorism calls on governments to apply the same level of detail and clarity in their legislation that they expect tech companies give to their users in publishing clear terms of service: clearly defined prohibitions, entrenched in the rule of law by aligning online and offline expectations of conduct, will avoid creating a differentiated regime for the online space

[126] https://www.techagainstterrorism.org/2021/02/17/position-paper-content-personalisation-and-the-online-dissemination-of-terrorist-and-violent-extremist-content/
[127] Rather than requiring platforms to assess content according to their own Terms of Service, or than requesting a court order to remove illegal content.
[128] *This is exemplified by the criticism made by David Kaye on the French "cyberhate law", see:* https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL_FRA_20.08.19.pdf

Legal Responses to Terrorist Use of the Internet: Africa

Few of the Commonwealth member countries in Africa have legal provisions specifically targeting use of the internet for terrorist purposes and the dissemination of terrorist content online. Sierra Leone, Nigeria and Kenya are notable exceptions, since their legislative frameworks include provisions specific to cyber-terrorism and terrorist content online. Tanzania's 2020 Regulations on Online Content also include provisions relating to terrorist content online, though not by prohibiting terrorist content specifically but by targeting statements related to terrorist acts.

In other African Commonwealth countries, existing legislation concerning the dissemination of terrorist material or public support for terrorism could be used to sanction terrorist content online as they do not specify the means of dissemination or support. Botswana, for instance, prohibits the creation and transmission of information relating to an act of terrorism, including documents available on the internet. Lawful limits to online content, even when not specifically targeting terrorist content, can potentially be engaged in the case of terrorist content. Malawi, for instance, provided for limitations to freedom of expression online for the purpose of protecting public order and national security in the Electronic Transactions and Cybersecurity Act 2017.

Below is an overview of how the law has responded to terrorist use of the internet and content in African Commonwealth member countries. The table summarises the main pieces of counter-terrorism legislation in the region, as well as other legislation concerning online content, and identifies the main provisions relating to countering terrorist presence online.[129]

Commonwealth Country	Relevant Law/Regulation/Body	Key takeaways
Botswana	Counter-terrorism Act (Act 24), 2014[130]	▪ The Counter-Terrorism Act does not specifically address terrorist use of the internet. ▪ However, the Act does consider as an offence associated with terrorism the act of "collect[ing], mak[ing], or transmit[ing] a record of information, or possess[ing] a document, record or thing" for the purpose of or in connection with an act of terrorism. ○ "transmitting" includes electronic transmission, or making an information / document available on the internet ○ This offence attracts a custodial sentence of up to 30 years in prison. ○ The Act makes no mention of liability for tech platforms.
The Gambia	Information and Communications Act of 2009[131]	▪ The Information and Communications Act does not specifically address terrorist use of the internet or content. ▪ However, the Act does penalise the use of computer systems to commit an offence under Gambian law.
Malawi	Electronic Transactions and Cybersecurity Act,	▪ Part IV of the Act concerns "Liability of Online Intermediaries and Content Editors and Protection of Online Users". ▪ The Act outlines limitations to the freedom of online communications, both to prohibit

[129] Please note: this table focuses solely on legal provisions that can be applied to online counter-terrorism efforts and thus excludes other laws aimed at regulating online content and the use of social media in a country such as the 2020 Proposed Amendments to the ICT Act for regulating the use and addressing the abuse and misuse of social media in Mauritius.
[130] https://issafrica.org/ctafrica/uploads/Botswana%20Counter-terrorism%20Act%202014.pdf; https://www.bocra.org.bw/bw-cirt;
[131] https://www.pura.gm/wp-content/uploads/2018/01/IC-Info-Comms-Act-2009.pdf

	July 2016 (came into force in 2017)[132]	"incitement of racial hatred, xenophobia or violence", and to "protect public order and national security, or enhance compliance with laws". o Given the nature of terrorist material, these prohibitions can be used to outlaw terrorist content shared online. ▪ The act states that intermediaries should not be liable for content shared on their services, unless they have initiated the transmission of the message, or had knowledge of the content, in which case they may be held liable. o The Act cautions that this should not be understood as a requirement to proactively monitor information. However, if a platform does monitor online communications and did not take action against prohibited content, it can be held liable. o Article 30 requires platforms to report to the Authority all illegal content uploaded by users, and to publicly disclose what they are doing to counter the dissemination of illegal content.
Mauritius	Prevention of terrorism Act, 2016[133]	▪ The Act states that a control order can be issued to prevent an individual from assessing certain technologies, including the internet, for the purpose of preventing terrorism.
Sierra Leone	Cyber Crime Act 2020, passed in June 2021[134]	▪ The Cyber Crime Act does in practice a regulate online content, o by outlining which content it is illegal to share in the country.135 o The prohibitions include online content of a terrorist, racist or xenophobic, and threatening nature. ▪ Article 41, on cyber terrorism, explicitly targets terrorist use of the internet" defined as "accessing or causing to be accessed a computer or computer system or network for purposes of a terrorist act".136 ▪ The Act announces the creation of a "National Cyber Security Incidence Response Coordination Center", responsible for responding to and managing cyber-incidents. This Center is to be headed by the National Cyber Security Coordinator and to include representatives from different ministers. ▪ The Act focuses, and thus places liability, on the person misusing computer systems rather than on tech platforms.

[132] https://malawilii.org/mw/legislation/act/2016/33;
https://www.justice.gov/eoir/page/file/1023986/download;

[133] https://www.ilo.org/dyn/natlex/docs/ELECTRONIC/104041/126736/F1839294808/MUS104041.pdf

[134] http://www.sierra-leone.org/Laws/2020-Cybercrime Act.pdf
https://www.switsalone.com/39316_parliament-of-sierra-leone-enacts-the-cyber-crime-bill/

[135] The Act specifies the purposes for which usage of a computer or computer system would illegal

[136] Terrorist acts are defined in the Anti-Money Laundering and Combating of Financing of Terrorism Act, 2012: http://www.sierra-leone.org/Laws/2012-02.pdf

		▪ The Act has been criticised by digital rights experts due to the surveillance power it grants to the government and law enforcement.[137]
Uganda	Uganda Communications Commission (UCC)[138]	▪ The UCC was created by the Communications Act of 2013, Section 5(1) to "among other things, monitor, inspect, license, set standards and enforce compliance relating to content".[139] ▪ Whilst the UCC does not currently regulate terrorist content, it has alluded to the possibility of doing so on different occasions: o In 2019 the UCC stated that it would focus on social media content and could possibly require online platforms to "filter/block/take down websites with specified content", including terrorism, hate speech, and incitement of violence.[140] o In an article about "regulation and responsible use of online media", the UCC mentioned the possibility of requiring ISPs to remove certain content including terrorist content, hate speech and incitement of violence.[141]
	Anti-terrorism Act, 2002[142]	▪ Prohibits the organisation of or support for acts promoting terrorism, as well as the publication and dissemination of related materials. Individuals found guilty of this can be sentenced to the death penalty.
Tanzania	The Electronic and Postal Communications (Online Content) Regulations, 2020[143]	▪ The Regulations prohibit various types of content,[144] including "Public security, violence and national safety", "criminal activities and illegal trade activities", "public information that may cause public havoc and disorder", and "false, untrue (misleading content)". o The Regulations do not explicitly prohibit terrorist content per se; however, it does prohibit the act of "[c]irculating information and statements with regard to possible terrorist attacks". ▪ The Regulations apply to users, ISPs, and to online content service providers (OSPs, including blogs and websites). ▪ The Regulations apply to all online content accessible by the public and "intended for consumption in or [having] originated from

[137] For instance, by requiring mobile operators to record voice calls and SMS in real time and to provide the recording to the authorities: https://www.mfwa.org/how-sierra-leon-is-hiding-behind-the-fight-against-cybercrime-to-abuse-digital-rights/

[138] https://www.ucc.co.ug/about-ucc/

[139]https://chapterfouruganda.org/resources/acts-bills/uganda-communications-act-2013#:~:text=The%20Act%20was%20passed%20to,to%20provide%20for%20related%20matters.

[140] https://freedomhouse.org/country/uganda/freedom-net/2020

[141] https://uccinfo.blog/2019/07/19/regulation-and-responsible-use-of-online-media/

[142] http://www.vertic.org/media/National Legislation/Uganda/UG_Anti-Terrorism_Act_2002.pdf

[143] https://thrdc.or.tz/wp-content/uploads/2019/09/ONLINE-CONTENT-REGULATIONS.pdf
The 2020 Online Content Regulations revise and revoke the EPOCA (Online Content) Regulations of 2018. See: https://www.article19.org/resources/tanzania-regulations-criminalise-free-speech/

[144] broadly more than its 2018 predecessor

		Tanzania". It does make special provision for content transmitted via private communications.[145] - Licensed ISPs and OSPs[146] are to immediately remove prohibited content when notified by the Tanzania Communications Regulatory Authority (TCRA); they are also required to proactively filter content and take corrective measures when necessary. - Non-licensed service providers (for instance foreign platforms or online content hosts) are still required to remove content upon notification by the TCRA or any affected parties, and to adopt a code of conduct for hosting content. - Platforms are potentially liable for failing to remove content, if flagged by the TCRA. – potential liability does not seem to apply for failing to remove content flagged by users. - With regard to user reporting, the Regulations states that "where an online content provider fails to resolve a complaint within twelve hours, the aggrieved person may, within thirty days from the date of filing the complaint, refer the complaint to the TCRA."
Kenya	The Computer Misuse and Cybercrimes Act, 2018[147]	- The Act establishes various offenses, including cyber terrorism, false publication of data, cyber harassment, identity theft and impersonation, and computer fraud.
	The Prevention of Terrorism Act (PTA), 2012[148]	- Kenya's legal framework to combat terrorism. Establishes terrorism-related offenses and provides the government with special investigative powers, as well as with special powers to arrest and detain suspects.
	The Security Laws Amendment Act, 2014[149]	- This Act amended the PTA to strengthen the country's counterterrorism efforts, and includes provisions on radicalisation and publishing offensive material. - The PTA and Security Laws Amendment Act enables national security bodies to intercept communications "for the purposes of detecting, deterring, and disrupting terrorism". The act also includes provisions on radicalisation as well as on the "publication of offending material".
Nigeria	The Cyber Crimes Act, 2015[150]	- Nigeria's framework for preventing and prosecuting cybercrimes in the country.

[145] https://www.africalegalnetwork.com/tanzania/wp-content/uploads/sites/23/2020/09/AK-Tanzania-The-Electronic-and-Postal-Communications-Online-Content-Regulations-2020.pdf

[146] The Regulations require OSPs to obtain a licence from the TCRA to provide "online content services", with different licences depending on the main types of content provided. The licence requirement does not appear to extend to foreign platforms, even though it is unclear how it will apply to foreign platforms with content accessible by the Tanzanian public.

[147] http://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/ComputerMisuseandCybercrimesActNo5of2018.pdf

[148] http://www.vertic.org/media/National Legislation/Kenya/KE_Prevention_Terrorism_Act.pdf

[149] http://kenyalaw.org/kl/fileadmin/pdfdownloads/AmendmentActs/2014/SecurityLaws_Amendment_Act_2014.pdf

[150] https://www.cert.gov.ng/ngcert/resources/CyberCrime__Prohibition_Prevention_etc__Act__2015.pdf

		▪ The Act prohibits the dissemination of various types of content and specifically penalises the use of the internet for terrorist purposes.
	Anti-terrorism Act, 2002[151]	▪ The act does not address terrorist use of the internet, but it does lay out the provisions for gathering intelligence from service providers and sharing information with other governments on terrorist use of the internet.

[151] http://www.vertic.org/media/National Legislation/Uganda/UG_Anti-Terrorism_Act_2002.pdf

Dossier C: Public-Private Partnerships and Cross-Industry Cooperation

Executive Summary

Since 2016, public-private partnerships aiming to tackle terrorist use of the internet have grown in number. Several initiatives focus on creating frameworks to more effectively limit terrorist content online. Such initiatives include the Christchurch Call to Action and the EU Internet Forum. Other initiatives, such as Internet Referral Units, are more operational in nature, and focus on notifying platforms of terrorist content.

Tech Against Terrorism is currently the only public-private partnership to focus on practical capacity building for smaller platforms. By acting as an intermediary between governments and tech platforms, the initiative has had a significant impact on the effort to tackle terrorist use of the internet. To date, Tech Against Terrorism has engaged with more than 400 global tech companies. It has a tangible record of success through its Mentorship Programme[152] and has also developed scalable products that help platforms improve their response to terrorist use of their services, such as the Terrorist Content Analytics Platform[153] and the Knowledge Sharing Platform.[154]

While most public-private partnerships focus on tackling terrorist content online, initiatives in other policy areas – particularly countering terrorist financing – can help inform online counter-terrorism efforts

Despite the increase in public-private partnerships over the last five years, there remain gaps in this joint effort which ought to be addressed to remedy the lack of:

- Activities focusing on practical impact, to include support mechanisms for smaller platform capacity building and targeted support for technical solutions
- Focus on smaller platforms in policy-driven strategic discussions
- Focus on the applicability of schemes to counter terrorist financing for online platforms

Recommendations

In line with and in addition to the gaps identified above, public-private partnerships should focus on:

- Promoting positive practical impact, particularly via support mechanisms for smaller platforms to increase resilience against terrorist use of their services.[155] This includes the development of data-driven tools to counter terrorist content online.[156]

- Ensuring smaller tech platforms are the focus of strategic policy discussions concerning terrorist use of the internet, including when considering updates to regulation

- Identifying opportunities to repurpose and adapt best practice in countering terrorist financing for online platforms

- Improving coordination and deduplication of effort. The focus of any given initiative, and the extent to which it may duplicate existing efforts, is not always clear. Policymakers should encourage coordination in public-private partnership activity in the online counter-terrorism area to maximise efficiency

[152] https://www.techagainstterrorism.org/2021/05/18/the-tech-against-terrorism-mentorship-2018-2021/
[153] https://www.terrorismanalytics.org/
[154] https://ksp.techagainstterrorism.org/
[155] For a summary of the threat faced by smaller platforms, see Dossier A.
[156] For a summary of Tech Against Terrorism's recommendations in this regard, see the gap analysis report produced by Tech Against Terrorism as co-chair of the GIFCT working group on technical approaches: https://gifct.org/wp-content/uploads/2021/07/GIFCT-TAWG-2021.pdf

Tech Against Terrorism: Public-Private Partnership

Inception and Mission

Tech Against Terrorism is an initiative launched in 2017 and supported by the United Nations Counter-Terrorism Executive Directorate (UN CTED). The official launch followed a first phase convened in April 2016, entitled 'Private Sector Engagement in Responding to the Use of the Internet and ICT for Terrorist Purposes: Strengthening Dialogue and Building Trust'.[157] Since 2019, Tech Against Terrorism has been implemented by the Online Harms Foundation.[158] The Tech Against Terrorism initiative operates pursuant to four UN Security Council Resolutions[159][160][161][162]as well as the Comprehensive International Framework to Counter Terrorist Narratives[163] that calls for improved public-private cooperation in tackling the use of the internet for terrorist purposes whilst respecting human rights. In its first year, Tech Against Terrorism worked closely with larger tech companies such as Facebook, Google, Microsoft, and Twitter, and in August 2017 supported their launch of the Global Internet Forum to Counter Terrorism (GIFCT). [164][165]

The objective of Tech Against Terrorism is to support the global tech sector in responding to terrorist use of the internet whilst respecting human rights. Tech Against Terrorism is a tech-agnostic initiative and works with companies across all types of technologies, with an explicit focus on supporting smaller tech companies with fewer resources to adequately address the urgent threat of terrorist exploitation. As a public-private partnership, Tech Against Terrorism works to foster constructive and improved working relationships between the tech sector and the public sector. Through its work, Tech Against Terrorism has directly engaged with more than 400 global tech companies.

Funding Model

Tech Against Terrorism is a public-private partnership funded by both the tech industry via the Global Internet Forum to Counter Terrorism (GIFCT) and by governments. To date, the governments of Spain, Republic of Korea, Switzerland, Canada, and the United Kingdom have provided financial support to Tech Against Terrorism.

Individual companies fund the delivery of specific bespoke services. The majority of Tech Against Terrorism funding is project-specific and tied to targeted deliverables associated with each funding body. Historically, funding provided by governments has been aimed at building products for tech companies to support their content moderation efforts. Examples of this include the Republic of Korea's funding for the initial version of Tech Against Terrorism's Knowledge Sharing Platform (see below), the United Kingdom's funding for the updated version, and the Canadian government's funding of the Terrorist Content Analytics Platform (TCAP – see below and Dossier F). In Tech Against

[157] A report on this phase can be downloaded here: https://www.techagainstterrorism.org/research/

[158] https://beta.companieshouse.gov.uk/company/11656320

[159] Resolution 2129 (2013) notes the evolving nexus between terrorism and the internet, and directs UN CTED to help address this

[160] Resolution 2354 (2017) mandates UN CTED to recommend ways for Member States to counter terrorist narratives

[161] Resolution 2395 (2017) recognises the development of Tech Against Terrorism and its efforts to foster collaboration between the tech industry, academia, and governments to disrupt terrorists' ability to use technology for terrorist purposes

[162] Resolution 2396 (2017) recognises the development of Tech Against Terrorism and its efforts to foster collaboration between industry, academia, and governments to disrupt terrorists' ability to use technology for terrorist purposes

[163] S/2017/375 Security Council proposal for a comprehensive international framework to counter terrorist narratives with a focus on public-private partnership - describing the Tech Against Terrorism initiative as good practice

[164] Global Internet Forum to Counter Terrorism to Hold First Meeting in San Francisco, Facebook, 13 July 2017 retrieved from https://newsroom.fb.com/news/2017/07/global-internet-forum-to-counter-terrorism-to-hold-first-meeting-in-san-francisco

[165] "Update on the Global Internet Forum to Counter Terrorism", Global Internet Forum to Counter Terrorism, 4 Dec 2017 retrieved from Facebook, YouTube, Microsoft, and Twitter: https://blog.twitter.com/official/en_us/topics/events/2017/GIFCTupdate.html, https://newsroom.fb.com/news/2017/12/update-on-the-global-internet-forum-to-counter-terrorism

Terrorism’s assessment this is one effective way to allow for government support for capacity building mechanisms for tech platforms via third parties

Public-Private Engagement and Coordination

Prior to the Covid-19 pandemic Tech Against Terrorism organised 15 global workshops which brought together tech companies, government entities, law enforcement agencies, civil society groups, and expert academics and researchers to identify ways to support smaller tech platforms. The workshops provided an unprecedented opportunity for smaller platforms to develop their awareness of the threat and to build constructive working relationships with stakeholders in the public sector. Via these workshops, of which a list appears below, Tech Against Terrorism has engaged with more than 160 global tech companies.[166]

Workshops organised by Tech Against Terrorism	
1 August 2017	San Francisco, United States
6 August 2017	Beirut, Lebanon
7 August 2017	Dublin, Ireland
18 September 2017	New York, United States
24 October 2017	Paris, France
30 October 2017	London, United Kingdom
7 November 2017	Jakarta, Indonesia
5 December 2017	Brussels, Belgium
7 May 2018	Abu Dhabi, United Arab Emirates
20 June 2018	Sydney, Australia
1 November 2018	Tel Aviv, Israel
7 December 2018	Berlin, Germany
18 June 2019	Amman, Jordan
14-15 November 2019	Delhi, India
10-11 December	London, United Kingdom

Tech Against Terrorism also facilitates coordination between platforms and governments, where relevant and appropriate. This includes facilitation of constructive working relationships. Where appropriate, Tech Against Terrorism acts as a coordinator between tech platforms and government agencies.

Work to Support Tech Platforms

Mentorship Programmes

The mentorship programme sits at the core of Tech Against Terrorism’s policy support work, helping tech companies to improve and future-proof their policies. Tech Against Terrorism also assists tech companies to develop the necessary processes and mechanisms to enforce their policies. All platforms participating in the mentorship programme benefit from Tech Against Terrorism’s expertise in counterterrorism and tech policy; Tech Against Terrorism strives to ensure that its support is as tailored as possible to a platform’s specificities and the threat of terrorist and violent extremist exploitation that it faces.

Tech Against Terrorism’s mentorship programme supports the Global Internet Forum to Counter Terrorism (GIFCT) and is designed to assist tech companies in meeting the Tech Against Terrorism and GIFCT Memberships criteria. Throughout the mentorship process, Tech Against Terrorism assists tech platforms in updating their policies and processes to meet the Tech Against Terrorism and GIFCT membership requirements.

Snapshot: Tech Against Terrorism Membership: Launched in 2017, the Tech Against Terrorism Membership is aimed at sharing best practice with tech platforms to help them build capacity in

[166] For more information about the workshops, see Tech Against Terrorism’s annual reports: https://www.techagainstterrorism.org/research/

tackling terrorist exploitation. This provides ongoing support following the mentorship process to ensure that we can continue assisting platforms as the threat landscape evolves. Tech Against Terrorism's Membership scheme is aimed at facilitating constructive working relationships built on trust with the global tech sector. Tech Against Terrorism's Membership is inclusive and we welcome tech companies of any size, region, technology, or service, offering to apply to become a member.

Snapshot: Global Internet Forum to Counter Terrorism Membership:
The GIFCT is an NGO designed to prevent terrorists and violent extremists from exploiting digital platforms. Founded by Facebook, Microsoft, Twitter, and YouTube as an industry coalition in 2017, and becoming an independent organisation in 2020, the Forum was established to foster technical collaboration among member companies, advance relevant research, and share knowledge with smaller platforms. The GIFCT launched its membership in 2019 based on the Tech Against Terrorism membership criteria. Tech Against Terrorism has been providing mentorship services for GIFCT applicants.

Tech Against Terrorism Membership has the following criteria:

1. Explicit prohibition of terrorism in Content Standards
2. Ability to both receive reports on content violations and act on them
3. Commitment to transparency reporting
4. A willingness to explore new technical solutions
5. A public commitment to respecting human rights, particularly freedom of expression and privacy
6. Support for civil society
7. Ability to receive user appeals and act on it (not necessary for GIFCT membership)

Membership in the GIFCT attracts one additional criterion:

8. Published transparency reports

Below we provide an overview of how Tech Against Terrorism supports tech platforms through the mentorship programme. For more information, please see the mentorship page on Tech Against Terrorism's website.[167]

In-depth policy review: As a first step in the mentorship process, Tech Against Terrorism conducts an in-depth review of the platform's content standards, highlighting their areas of strength and outlining where there is room for improvement. Tech Against Terrorism examines all publicly available company policies relating to content moderation and counterterrorism. This review allows for a precise understanding of a platform's approach to online moderation and ensures that Tech Against Terrorism can provide bespoke support.

Bespoke policy recommendations: Based on the policy review and informed by Tech Against Terrorism's expertise and research, it provides bespoke policy recommendations for all Tech Against Terrorism mentees. These recommendations focus on ensuring that platforms have policies in place to adequately counter terrorist and violent extremist exploitation of their services. Tech Against Terrorism's recommendations also aim to ensure that policies are capable of being operationalised and adapted to meet the threat each platform faces.

Tech Against Terrorism recommendations to its members are based on the following principles:

- Platforms should have clear and detailed language prohibiting terrorism or violent extremism in their content standards.
- The prohibition of terrorism should be compliant with the rule of law, for example by referencing a terrorist designation list (such as the UN Security Council Consolidated List). As a basis for a publicly available list of terrorist organisations, tech companies can also

[167] https://www.techagainstterrorism.org/2021/05/18/the-tech-against-terrorism-mentorship-2018-2021/

refer to the Group Inclusion Policy developed for the Terrorist Content Analytics Platform (TCAP).

- Content moderation practices should respect human rights and freedom of expression, in line with the Tech Against Terrorism Pledge.
- Platforms should have clear and detailed community guidelines so that users know what is expected of them, and clearly explained moderation processes so that users know what they can expect of the platforms.
- Platforms should strive to develop dedicated counterterrorism and violent extremism policy to demonstrating that the threat has been considered and will be dealt with seriously.
- Platforms should have user-initiated reporting mechanisms in place to ensure that content and behaviour violative of expected standards can be reported by users
- Platforms should have user appeal processes in place to ensure that users can seek redress if they believe their content was wrongfully actioned.

Transparency & accountability: Tech Against Terrorism's support programme is underpinned by its concomitant increase to meaningful transparency and accountability from the tech sector. Tech Against Terrorism focusses on helping platforms to draft clear and precise policies and content moderation rules so that users know what is expected of them and what the platform is doing to ensure safety on its services. All of Tech Against Terrorism's policy reviews also include a section dedicated to transparency and transparency reports. Tech Against Terrorism assesses and provides recommendations on a platforms' general transparency efforts and, for mentees who have already published a transparency report, conducts a complete assessment of the most recent transparency reports available.

Providing support beyond policy recommendations: Tech Against Terrorism's policy support extends beyond policy recommendations to ensure that platforms have the necessary tools to implement them. Tech Against Terrorism provides mentees with various resources on policy best practices, and with advisory support to reviewing policy updates.

Knowledge-sharing: The Mentorship Programme sits at the core of Tech Against Terrorism's knowledge-sharing activities. Tech Against Terrorism works with a broad range of stakeholders to build on its understanding of the threat and of the evolution in counter-terrorism practices. The multiple different insights into policy and best practice which Tech Against Terrorism gathers from its extensive networks are used to inform its support of tech companies. All mentees benefit from privileged access to Tech Against Terrorism's different knowledge-sharing activities, including its e-learning webinar series in collaboration with the Global Internet Forum to Counter Terrorism (GIFCT), and Tech Against Terrorism's own resources on countering terrorist use of the internet and understanding the threat. The updated Knowledge-Sharing Platform (KSP) also supports this work (see below).

OSINT monitoring and intelligence briefs: All mentees benefit from Tech Against Terrorism's open-source intelligence (OSINT) capacity. Each mentee receives a bespoke intelligence brief providing an overview of the terrorist and violent extremist threat to its platform. Tech Against Terrorism's analysis of the terrorist and violent extremist online landscape also informs its understanding of the threat and support to platforms.

Continuing support and engagement: Throughout the mentorship and membership process Tech Against Terrorism strives to provide regular support to its mentees. Each mentee has a dedicated Tech Against Terrorism point of contact responsible for overseeing the mentorship process and responding to any questions that might arise regarding the mentorship process, or regarding terrorist use of the internet generally. Tech Against Terrorism also organises regular catch-up calls to ensure that Tech Against Terrorism is aware of any challenges that might arise for its mentees.

8 of our mentees have completed their mentorship and **joined the Tech Against Terrorism Membership**

8 of our mentees have already **been accepted as GIFCT Members**

19 of our mentees have already **signed the Tech Against Terrorism Pledge**

10 of our mentees have already **updated their content standards based on our recommendations**

5 of our mentees have already **strengthened their approach to counterterrorism based on our advice**

4 of our mentees have already **published a transparency report with our support, 1 has updated its reporting metrics to cover T/VE content**

6 of our mentees are currently **working on producing their first transparency reports with our help**

Terrorist Content Analytics Platform

Launched in November 2020, **The Terrorist Content Analytics Platform**[1] (TCAP), developed by Tech Against Terrorism, constitutes the world's largest database of terrorist content collected in real time from verified terrorist channels on messaging platforms and apps. This repository of verified terrorist content (imagery, video, PDFs, URLs, audio) collected from open sources and existing datasets facilitates secure intelligence sharing between platforms.

Developed support from Public Safety Canada,[168] TCAP functions as a secure online tool automated to detect and analyse verified terrorist content on smaller internet platforms. Following detection, the TCAP alerts tech companies to terrorist content located on their platforms and supports smaller platforms to improve their content moderation. The TCAP will also improve academic research on terrorist content and augment efforts to use artificial intelligence (AI) and machine learning to detect terrorist content at scale.

Since November 2020, the TCAP has sent 7,400 alerts to 59 different tech companies. For more information about the TCAP, please see Dossier F.

Knowledge Sharing Platform

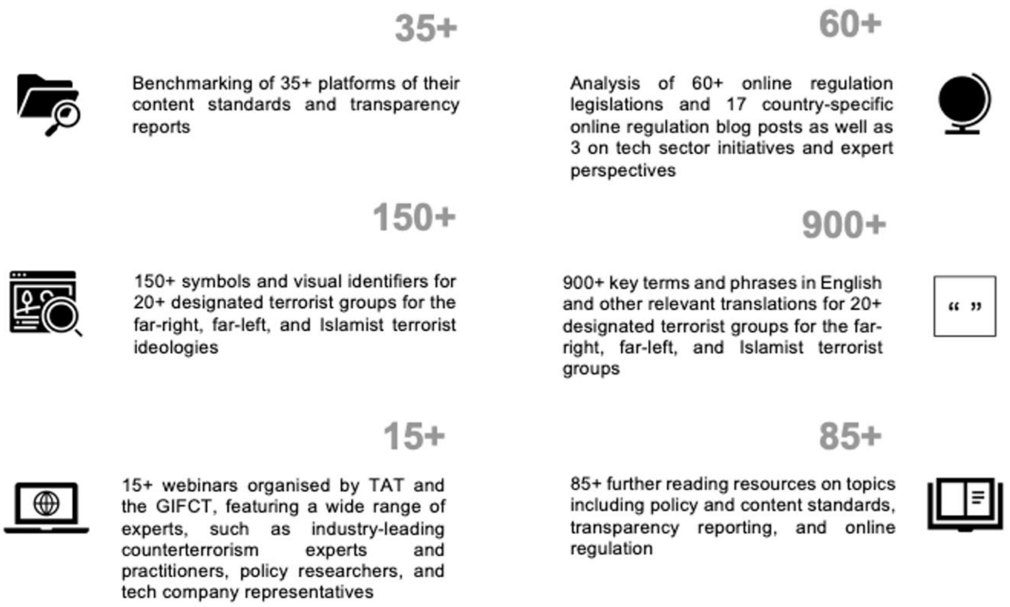
The Knowledge0Sharing Platform (KSP) is a platform developed by Tech Against Terrorism to provide smaller tech companies with a collection of interactive tools and resources to support the creation of an effective and human rights compliant counter-terrorism response.

The KSP is a free platform which provides content moderators and Trust & Safety teams with information, guidelines, and recommendations to support them in tackling terrorist and violent extremist use of their platforms whilst also increasing transparency and accountability towards their users.

Tech Against Terrorism works with a broad range of stakeholders to build on its understanding of the threat and evolving counter-terrorism practices. The different insights and best practice Tech Against Terrorism gathers from its extensive networks are used to inform its policy and practical support of tech companies. The KSP's resources include research and analysis on terrorist use of the internet, such as the threat landscape and proscribed organisations; on global online regulation; as

[168] https://www.techagainstterrorism.org/2019/06/27/press-release-tech-against-terrorism-awarded-grant-by-the-government-of-canada-to-build-terrorist-content-analytics-platform/

well as guidelines and recommendations on content standards and transparency reporting. Resources include:



35+

Benchmarking of 35+ platforms of their content standards and transparency reports

60+

Analysis of 60+ online regulation legislations and 17 country-specific online regulation blog posts as well as 3 on tech sector initiatives and expert perspectives

150+

150+ symbols and visual identifiers for 20+ designated terrorist groups for the far-right, far-left, and Islamist terrorist ideologies

900+

900+ key terms and phrases in English and other relevant translations for 20+ designated terrorist groups for the far-right, far-left, and Islamist terrorist groups

15+

15+ webinars organised by TAT and the GIFCT, featuring a wide range of experts, such as industry-leading counterterrorism experts and practitioners, policy researchers, and tech company representatives

85+

85+ further reading resources on topics including policy and content standards, transparency reporting, and online regulation

The KSP resources are meant to alleviate the burden on smaller platforms' content moderators and Trust & Safety teams in responding to terrorist content and other illegal activity. Although there are many resources available online which promote understanding of the terrorist threat landscape, online regulation, and human rights, such resources are difficult for tech companies to locate and often accompanied by little analysis. The KSP collates and organises all this information, and further provides analysis and actionable guidelines in a hub of learning which contains operable policy and enforcement recommendations for smaller platforms.

Bespoke Services

Tech Against Terrorism provides bespoke services for companies at a cost. These services can include:

Type	Example
Research and Analysis	Mapping and analysis of historical terrorist and violent extremist use of the platform and other similar platforms
	Real-time Open Source Intelligence (OSINT) monitoring
	Bespoke research support based on company requirements
Threat intelligence and risk assessments	Future-oriented threat intelligence and risk assessments
	Red teaming and usage simulation
Mentorship and policy guidance	Mentorship and policy guidance
	Practical support, including in developing transparency reporting or in updating content standards
Crisis coordination response	Crisis coordination support, including OSINT monitoring and support in suspected threat to life situations
Data-driven solutions support	Data science modelling support
	Classifier / tool creation & partnership with 3rd parties

Work to Support Governments

Confidential Threat Assessment Briefs

Tech Against Terrorism's OSINT team regularly produces confidential threat intelligence and assessment briefs for government stakeholders. These briefs provide an update on how terrorist actors use online platforms and caution against future risks. If your government is interested in receiving these briefs, please get in contact.

Contribution to Government Led Processes

Tech Against Terrorism regularly contributes to processes led by governmental and inter-governmental organisations. Since its inception in 2017, Tech Against Terrorism has participated in more than 120 conferences in 29 countries.[169] It regularly contribute to consultation processes for government-led initiatives, including new legislation pertaining to terrorism and online regulation. Examples of the latter include the European Union's Digital Services Act,[170] Australia's Online Safety Act,[171] and the United Kingdom's regulation on video-sharing platforms.[172]

Existing public-private partnerships aimed at tackling terrorist use of the internet

Multistakeholder processes focussed on online content

Christchurch Call to Action

The Christchurch Call to Action[173] is a commitment by Governments and tech companies to counter terrorist and violent extremist content online. The Call was initiated by New Zealand Prime Minister Jacinda Ardern and French President Emmanuel Macron in 2019 in response to the terrorist attack on 15 March 2019 in Christchurch, New Zealand. Supporters of the Call to Action include governments, tech companies, and initiatives such as the Global Internet Forum to Counter Terrorism, the EU Internet Forum, the G20, the G7, and the Aqaba Process.[174]

The Call outlines 25 commitments – five for governments, seven for tech companies, and 13 for both sectors – that signatory governments and tech companies agree to work towards. The commitments call on governments and tech companies to work to prevent terrorist use of the internet and to tackle the root causes of terrorism in line with human rights and fundamental freedoms safeguards.[175][176][177]

The Christchurch Call Advisory Network[178]is a network comprised of 47 organisations aiming to tackle terrorism online whilst maintaining a free, open, and secure internet and without compromising human rights and fundamental freedoms.[179] The Christchurch Call Advisory Network

[169] For more information about specific conferences attended, please see Tech Against Terrorism's annual reports: https://www.techagainstterrorism.org/research/

[170] https://www.techagainstterrorism.org/2020/09/29/summary-tech-against-terrorisms-response-to-the-eu-digital-services-act-consultation-process/

[171] https://www.techagainstterrorism.org/2021/02/18/tech-against-terrorism-summary-of-consultation-response-regarding-the-government-of-australias-online-safety-bill/

[172] https://www.techagainstterrorism.org/2021/07/02/tech-against-terrorism-submission-to-the-consultation-process-for-ofcoms-consultation-on-guidance-for-vsp-providers-on-measures-to-protect-users-from-harmful-material/

[173] https://www.christchurchcall.com/call.html

[174]https://www.techagainstterrorism.org/2021/05/14/statement-on-the-second-anniversary-of-the-christchurch-call-to-action/

[175] https://www.christchurchcall.com/call.html

[176] The Call received its name from the 15 March 2019 Christchurch mosque attacks where a far-right terrorist killed 51 people in the city of Christchurch, New Zealand. The attacker spread his manifesto online and live-streamed his attack, and a recording subsequently went viral across social media platforms – our summary of how tech platforms were used to spread this content, and how they responded, is available at: https://www.techagainstterrorism.org/2021/05/14/statement-on-the-second-anniversary-of-the-christchurch-call-to-action/

[177] https://www.christchurchcall.com/christchurch-call.pdf

[178] https://www.christchurchcall.com/advisory-network.html

[179]https://www.techagainstterrorism.org/2021/05/14/statement-on-the-second-anniversary-of-the-christchurch-call-to-action/

membership is drawn from several civil society groups concerned variously with human rights, freedom of expression, digital rights, counter-radicalization, victim support and public policy.[180]

EU Internet Forum

The EU Internet Forum (EUIF) is a public-private multi-stakeholder initiative intended to tackle terrorist use of the internet in the EU. The EUIF was created in 2015 and convenes Member States, tech companies, and relevant expert stakeholders[20] with the aim of creating joint voluntary approaches to preventing terrorist use of the internet and hate speech. Since 2020, child sexual exploitation has also been included in discussions at the Forum. Meetings are usually divided into separate technical and ministerial components, and tech companies like Facebook, Google, Twitter, and Telegram regularly attend the meetings.

EU Code of Conduct Against Hate Speech

The EU Code of Conduct on Illegal Hate Speech is a voluntary collaborative protocol between the EU and tech companies, in which platforms undertake to remove and report on illegal hate speech, as outlined in the EU's Framework Decision 2008/913/JHA,[181] notified by users, law enforcement, and a select number of European civil society groups.[182] Participant companies are expected to have in place policies prohibiting – and processes to review notifications regarding – illegal hate speech. Upon receipt of a removal notification, companies review the content against their content standards and – where necessary – national laws transposing the Framework Decision 2008/913/JHA. Companies are expected to review the majority of notifications in less than 24 hours and remove or disable access to such content, if found to violate their policies and/or relevant laws. Between 2019-2020, 4,364 pieces of content were flagged to participating companies, 94% of which were reviewed by the companies within 24 hours, and 71% of which were removed.[183]

There exists some criticism against the EU-led PPPs. Voluntary arrangements like EUIF and the Code of Conduct have been criticised for setting undue speech regulation under the guise of volunteerism. Professor Danielle Citron described the EUIF as an example of the EU contributing to 'censorship creep'.[21] According to Citron, several of the voluntary steps that tech companies have taken to address terrorist use of their platforms since 2015 have been made specifically to placate EU legislators. Whilst Citron acknowledges that results have come out of this approach (the GIFCT hash-sharing database – see Dossier D – is one example), the definitional uncertainty around terms like terrorist content means that there is significant risk of erroneous removal, which negatively impacts freedom of expression. Further, since companies are tackling content "voluntarily", material is removed under company speech policies rather than local or regional legislation, meaning that effects are global effects despite being based on European standards.

Internet Referral Units

Internet Referral Units are dedicated bodies working to discover and refer terrorist content to tech companies, who then assess content against their own internal rules and policies. Engagement is voluntary and tech companies are not formally obliged to remove content referred to them.

United Kingdom: The UK Counter-terrorism Internet Referral Unit (CT IRU) detects and refers terrorist content to tech platforms. The first IRU in Europe, the Counter Terrorism Internet Referral Unit (CTIRU), was established in the UK in 2010, and is considered to have spurred the creation of other IRUs in Europe.[184] The CTIRU was set up by the National Police Chiefs' Council and is run by the UK Metropolitan Police, with a legal mandate from the UK Terrorism Act (2006). The CTIRU refers content to tech platforms based official assessment of whether content meets the definition of terrorist material provided in the UK legislation, but there is currently no legal instrument that requires companies to remove the content. However, referrals would, under the EU

[180] https://www.christchurchcall.com/advisory-network.html
[181] https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32008F0913
[182] https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en; https://ec.europa.eu/info/sites/default/files/codeofconduct_2020_factsheet_12.pdf.
[183] https://ec.europa.eu/info/sites/default/files/codeofconduct_2020_factsheet_12.pdf
[184] https://globalnetworkinitiative.org/human-rights-risks-irus-eu/

e-Commerce Directive by which the UK continues to abide, count as actual knowledge of the content capable of establishing liability for not removing it. It is therefore likely that content will be removed. Since the CTIRU does not publish transparency reports there is no definitive account of the scale of their operations, however a UK Government spokesperson in 2019 said that more than 310,000 pieces of terrorist content had been removed from the web as a result of the CTIRU's work.[185] Civil society organisation Open Rights Group has also collated publicly available sources on CTIRU removal statistics.[186]

The EU Internet Referral Unit: The EU Internet Referral Unit is based on the model pioneered by the UK's CTIRU. The EU IRU employs subject matter experts to refer suspected Islamist terrorist content to tech companies, who then assess whether the content violates their Terms of Service. Member States are also able to refer content to the EU IRU. The unit conducts so-called referral assessment days with tech companies. This has led to substantial removal of terrorist content, including a joint operation with Telegram[187] to remove a large number of Islamic State channels. According to the EU IRU, the Unit has to date referred more than 111,000 pieces of content[188] to tech companies.

France: France's IRU, the *Office central contre la criminalité liée aux technologies de l'information et de la communication* (OCLCTIC)[189] refers content to tech platforms in line with Law 2014-1353 on Strengthening Provisions to Fight Terrorism. The OCLCTIC sits under the Ministry of Interior, and it is unclear if it is by reference to a company's Terms of Service or to statutory authority that content is reviewed and actioned .[190]

The Netherlands: The Dutch IRU refers content to tech companies after assessing it against Dutch law under the Police Act of 1993. There does not seem to be any further public information about their activities.

There exists criticism against IRUs. Civil society and digital rights groups have long criticised IRUs for their lack of transparency. The Global Network Initiative (GNI) has noted that there is no formal oversight of judicial review of IRU activities. Further, experts have also pointed out that IRUs contribute to content removal via extra-legal means, in the sense that it is a government-affiliated body which is encouraging content removal but without using legal orders.

Counter Terrorism Financing PPPs

Below, we summarise noteworthy public-private partnerships seeking to tackle terrorist financing and the extent to which they engage in CTF practices relevant to the online sphere.[191] Whilst not specifically designed to tackle terrorist use of the internet, there are several examples of public-private collaboration from the areas of anti-money laundering and countering terrorist financing that may provide useful best practice.

Examples of global countering terrorist financing regimes and bodies include:

Financial Action Task Force (FATF): The FATF is an inter-governmental organisation founded in 1989 to combat money laundering. In 2001, its mandate was widened to include countering terrorist financing. The FATF has 39 member jurisdictions (37 countries as well as the EU Commission and the Gulf Cooperation Council). There are also nine associate members, which each represent regional FATF style bodies. There are three African associate members which see Commonwealth states amongst its membership:

[185] https://www.theyworkforyou.com/wrans/?id=2019-06-26.269681.h
[186] https://wiki.openrightsgroup.org/wiki/Counter-Terrorism_Internet_Referral_Unit#cite_note-52
[187] https://www.europol.europa.eu/newsroom/news/referral-action-day-against-islamic-state-online-terrorist-propaganda
[188] https://www.europol.europa.eu/publications-documents/eu-iru-transparency-report-2019
[189] https://www.police-nationale.interieur.gouv.fr/Organisation/Direction-Centrale-de-la-Police-Judiciaire/Lutte-contre-la-criminalite-organisee/Sous-direction-de-lutte-contre-la-cybercriminalite
[190] https://globalnetworkinitiative.org/human-rights-risks-irus-eu/
[191] For a summary of how terrorists use the internet for financing purposes, please see dossier A. A summary can also be found in the Asia Pacific Group and the Middle East and North Africa Financial Action Task Force's 2019 report: http://www.apgml.org/methods-and-trends/news/details.aspx?pcPage=1&n=1142

Inter-Governmental Action Group against Money Laundering in West Africa (GIABA): GIABA is a specialized institution of Economic Community of West African States (ECOWAS) that is responsible for strengthening the capacity of member states towards the prevention and control of money laundering and terrorist financing in the region. It is also a FATF-Styled Regional Body (FSRB) working with its member States to ensure compliance with international AML/CFT standards. GIABA's membership consists of 16 Member States, including Commonwealth members Ghana, Nigeria, the Gambia, and Sierra Leone. As part of its research outputs GIABA in 2020 examined risks and opportunities for CTF activities as a result of increasing use of fintech in the West Africa Region.

Eastern and Southern Africa Anti-Money Laundering Group (ESAAMLG): ESAAMLG is a Regional Body subscribing to global standards to combat money laundering and financing of terrorism and proliferation. Amongst its 18 Member States are Commonwealth countries Botswana, Eswatini, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Rwanda, Seychelles, South Africa, Tanzania, Uganda, and Zambia. The ESAAMLG describes its mission as combating money laundering by implementing the FATF Recommendations. This effort includes coordinating with other international organisations concerned with combating money laundering, studying emerging regional typologies, developing institutional and human resource capacities to deal with these issues, and coordinating technical assistance.

Task Force on Money Laundering in Central Africa (*Groupe d'Action contre le blanchiment d'Argent en Afrique Centrale* (GABAC)): GABAC was established in 2000 with the mandate to combat money laundering and terrorist financing, assess the compliance of its members against the FATF Standards, provide technical assistance to its member States and facilitate international co-operation. It currently consists of seven member states, of which Commonwealth member Cameroon is one. It also sees membership from the Governor of the Banks of the States of Central Africa, the President of the Monetary Community of Central Africa (CEMAC) Commission, the President of the Committee of Police Chiefs of Central Africa, and Secretary General of the Banking Commission of Central Africa.

United Kingdom: Joint Money Laundering Intelligence Task Force (JMLIT): JMLIT is a partnership between law enforcement and the financial sector to exchange and analyse information relating to money laundering and wider economic threats. According to the UK Government, the taskforce consists of "over 40 financial institutions" as well the country's Financial Conduct Authority, non-profit fraud prevention organisation Cifas, the National Crime Agency, Her Majesty's Revenue and Customs (HMRC), the City of London Police, the Metropolitan Police Service, and the Serious Fraud Office. Since its inception in 2015, JMLIT has supported "over 500 law enforcement investigations which ha[ve] directly contributed to over 130 arrests and the seizure or restraint of over £13 million."[192] In addition to disrupting financial crime, including terrorist financing, the members also work together to enhance their mitigation strategies.[193] JMLIT is structured across an Operational Group, multiple Expert Working Groups and an Alerts Service for the wider dissemination of assessments and typologies, which is provided by UK Finance. Activities are overseen by a Management Board, which reports to the Financial Sector Forum, a body that facilitates high-level dialogue between the financial sector, the National Crime Agency (NCA) and the Financial Conduct Authority (FCA) and is overseen by the UK Home Office. The Operational Group "brings together dedicated vetted representatives of the large retail and investment banks, law enforcement agencies and the FCA to share information on operational-level activity."[194]

Australia: Fintel Alliance: Fintel Alliance was established in 2017 by the Australian Transaction Reports and Analysis Centre (AUSTRAC), which describes it as a "world first public-private partnership". Its partners include banks, remittance service providers, gambling operators, and law enforcement agencies.[195][196] Fintel Alliance produces "financial crime guides" to raise awareness of money laundering, terrorist financing and other unlawful financing amongst businesses.[197] The body also works with AUSTRAC to investigate and disrupt terrorist activity. Fintel Alliance has also

[192] https://www.nationalcrimeagency.gov.uk/what-we-do/national-economic-crime-centre

[193] https://www.nationalcrimeagency.gov.uk/what-we-do/national-economic-crime-centre

[194] https://www.gov.uk/government/news/anti-money-laundering-taskforce-unveiled

[195] https://www.austrac.gov.au/about-us/fintel-alliance

[196] https://www.future-fis.com/uploads/3/7/9/4/3794525/ffis_report_-_oct_2017_web.pdf

[197] https://www.austrac.gov.au/business/how-comply-and-report-guidance-and-resources/guidance-resources/all-resources?field_industries_target_id=All&field_guidance_topics_target_id=All&field_resource_type_target_id=87

established an "innovation hub", which assesses the impact of emerging technologies like blockchain and cryptocurrencies on money laundering and terrorist financing.[198] The body has noted the risk of "increased misuse of online financial services" in one of its performance reports.[199]

Singapore: Anti-Money Laundering and Countering the Financing of Terrorism Industry Partnership (ACIP): ACIP was founded in 2017 and brings together Singapore's financial sector, regulators, law enforcement and government agencies.[200] ACIP also tasks expert working groups to study specific topics for the benefit of the body. It also releases "best practice papers" for financial institutions to help protect them against criminal activity, including terrorism.[201]

Targeted project or workstream initiatives

Crisis protocols

EU Crisis Protocol: In 2019, the EU Internet Forum committed to creating a crisis protocol to prevent viral spread of terrorist material in the immediate aftermath of a terrorist attack. The protocol was created in response to the Christchurch Call to Action. The protocol is a voluntary mechanism by which governments and tech companies commit to identify, notify, and share information about terrorist content that risks becoming viral. All contributing parties have an assigned point of contact. In the event of a potential "crisis" (defined as "where terrorist and violent extremist content spreads online rapidly"), the protocol asks contributing partners to take target (attack location, number of platforms associated content is found on, attack type, and victims affected) and impact (content virality, reproducibility, and resilience) into account to assess whether an event meets the crisis threshold. Based on that assessment, parties notify and share information to prevent content virality. Post-crisis reports are also produced and shared between contributing partners.[202]

Christchurch Call Shared Crisis Response Protocol: One of the priorities of the Christchurch Call to Action was to develop a shared crisis protocol to allow for improved information sharing between governments and tech companies in the event of a crisis. While there is not much publicly available information on how the protocol works, in December 2019 the protocol was reviewed by representatives of government, the tech industry, and civil society; the recently published Christchurch Call to Action Crisis Response Workstream lays out the priorities for developing this over the next year.

Transparency reporting

The Organisation for Economic Cooperation and Development (OECD) in 2019 launched a workstream to develop a voluntary framework to encourage tech platforms to increase transparency of actions taken to counter terrorist and violent extremist content on their platforms. The workstream is sponsored by the governments of Australia and New Zealand, and convenes stakeholders from governments, tech companies, civil society groups and academia. Tech Against Terrorism has participated in the process since its inception. The voluntary framework is intended to consist of inclusion metrics that signatory companies would be expected to meet. The workstream was initiated at a workshop in Boston in November 2019, but has been continued virtually throughout 2020 and 2021 due to the Covid-19 pandemic.[203] It is unclear when the project will finish and what support mechanisms will be provided for tech companies to be able to meet the voluntary requirements.

Digital evidence

Practical guide for requesting electronic evidence across borders: In 2018, the United Nations Counter Terrorism Executive Directorate (UN CTED), the United Nations Office on Drugs and Crime (UNODC) and the International Association of Prosecutors (IAP) initiated a multistakeholder process to develop a practical guide to requesting digital evidence across borders to benefit investigations, including

[198] https://www.austrac.gov.au/about-us/fintel-alliance
[199] https://www.austrac.gov.au/sites/default/files/2020-11/Fintel%20Performance%20Report%202020.pdf
[200] https://www.future-fis.com/uploads/3/7/9/4/3794525/ffis_report_-_oct_2017_web.pdf
[201] https://www.mas.gov.sg/regulation/anti-money-laundering/amlcft-industry-partnership-acip
[202] https://ec.europa.eu/home-affairs/sites/default/files/what-we-do/policies/european-agenda-security/20191007_agenda-security-factsheet-eu-crisis-protocol_en.pdf
[203] https://oecd-innovation-blog.com/2020/09/15/terrorist-violent-extremist-content-internet-social-media-transparency-tvec/.

investigations of terrorist activity.[204] The Guide was developed through several global workshops with stakeholders from law enforcement, government, the tech sector, and civil society,[205] and includes practical guidance on how law enforcement agencies can best request evidence from tech platforms in line with said platforms' policies and processes. The Guide was published in 2019 and is accessible through UNODC's SHERLOC portal.[206]

Gap analysis: public-private partnerships and terrorist use of the internet

Based on this analysis, Tech Against Terrorism identifies the below as critical gaps currently unaddressed by the existing global public-private partnerships we have examined. Our findings are indicative and are not intended as a definitive summary.

- Lack of focus on smaller platforms in policy-driven strategic discussions
- Lack of support mechanisms for capacity building by smaller platforms, including targeted support for technical solutions
- Lack of focus on the applicability for online platforms of schemes to counter terrorist financing schemes

Lack of consideration of smaller platforms in policy-driven strategic discussions

Whilst the initiatives mentioned above have been successful at convening government and private sector stakeholders – and whilst consideration for smaller tech companies has improved – many of the collaborative schemes place a disproportionate focus on larger tech companies.[207] Whilst this is not entirely misplaced and whilst the above initiatives have clearly led to successes, smaller tech companies need to be front and centre of any initiative seeking to counter terrorist use of the internet due to the threat of terrorist and violent extremist actors exploiting such services. This is particularly pressing in some of the strategic discussions that take place in public-private partnership forums, which often focus on encouraging improved action and concrete action from larger tech companies or industry bodies such as the GIFCT. It is Tech Against Terrorism's assessment that government agencies are aware on a technical level of the threat that exists on smaller platforms[208] and that this is reflected in strategic policy discussions.

Lack of practical support mechanisms for smaller platforms

There is generally a lack of practical support mechanisms offered to smaller tech platforms from public-private partnerships. Whilst there are commendable exceptions, in Tech Against Terrorism's assessment many of the initiatives place disproportionate emphasis on engagement with larger tech companies or on working towards frameworks and standards that place expectations on all tech companies but without a concrete plan as to how smaller tech companies will be supported in meeting such expectations. Currently, Tech Against Terrorism is the only initiative to focus on this.

Gaps in technical approaches to countering terrorist use of the internet

There are gaps in technical capacity building initiatives. In Tech Against Terrorism's gap analysis report on technical approaches (including automated, data-driven machine-learning and artificial intelligence solutions), published in July 2021,[209] Tech Against Terrorism examined which areas should be prioritised. Below is a summary of some of its key concerns:

[204] https://www.unodc.org/unodc/en/frontpage/2018/February/experts-meet-in-vienna--discuss-lawful-access-to-digital-data-across-borders.html.

[205] Tech Against Terrorism participated in three of these workshops.

[206] https://www.unodc.org/unodc/frontpage/2020/October/unodc-promotes-international-cooperation-in-sharing-electronic-evidence-with-global-partners.html

[207] Some of the collaborative initiatives have however incorporated smaller platforms in their response. For example, the EU Internet Forum has for several years seen participation from a wide range of tech companies, including micro-platforms. Likewise, the Christchurch Call mentions smaller platforms and the need to provide capacity building for such actors.

[208] For example, transparency reports by Europol clearly demonstrate that the majority of online activity from groups like IS and al-Qaeda occurs on smaller platforms as opposed to the large social media companies. See: https://www.europol.europa.eu/newsroom/news/eu-iru-transparency-report-2019

[209] https://gifct.org/wp-content/uploads/2021/07/GIFCT-TAWG-2021.pdf

- Lack of focus on solutions beyond content removal, including on content moderation workflow solutions and the underlying processes. This is arguably driven by a an over-zealous (and misplaced) belief on the part of policymakers that technology can solve the challenges of terrorism on its own without attention to associated difficulties of policy.
- Lack of targeted support mechanisms for smaller priority platforms to facilitate and coordinate implementation of tech solutions across at-risk platforms. This is arguably informed by a limited understanding of the current threat picture with regard to terrorist use of the internet and the constraints in capacity and capability faced by smaller platforms.

Lack of adaptation of countering terrorist financing schemes to online platforms

There are many useful lessons to be drawn from public-private partnerships concerned with countering terrorist financing. However, the relevance of how such lessons can be implemented, adapted, or repurposed for online platforms requires attention. Such platforms include fintech, crowdfunding, payment apps, and cryptocurrencies, but social media and other content sharing platforms should also be included.[210][211] This includes thinking through how to strike a good balance, as noted by a 2020 GIABA report on fintech, between the stringent know-your-customer practices observed in the financial sector and the inclusiveness of fintech platforms.[212]

[210] https://rusi.org/explore-our-research/publications/special-resources/social-media-and-terrorist-financing-what-are-vulnerabilities-and-how-could-public-and-private

[211] http://www.apgml.org/methods-and-trends/news/details.aspx?pcPage=1&n=1142

[212] https://www.giaba.org/media/f/1125_ENG%20-%20ML-TF%20RISKS%20AND%20OPPORTUNITIES%20OF%20FINTECH-3.pdf

Dossier D: Tech Sector Efforts to Counter Terrorism and Cross-Industry Cooperation

Executive Summary

- Collaborative industry efforts such as the Global Internet Forum to Counter Terrorism (GIFCT), the Trust & Safety Professional Association (TSPA), and the Trust & Safety Foundation (TSF) aim to provide practical capacity building and knowledge sharing for tech companies. The GIFCT facilitates collaboration and knowledge sharing amongst the tech sector to tackle terrorist use of the internet. By contrast, the TSPA and its sister organisation, the TSF, enforce principles and policies that define acceptable behaviour and content online. They do so by providing a space for professionals in the field and focussing on improving society's understanding of trust and safety.

- Tech sector efforts to counter terrorism encompass a range of methods, including counter-terrorism policies and robust content standards, content moderation identification and solutions, and transparency reporting. In this report, we analyse the overarching trends in tech sector efforts, as well as assessing the policies and practices of local platforms.

- Companies engaged in preventing terrorist or violent extremist exploitation of their platform do so by multiple means, which require a range of resources such as people, policies, processes, and systems. This includes prohibiting content in their Terms of Service or Community Guidelines, and other relevant content standards; identifying and detecting prohibited content, such as through user reporting or proactive monitoring; providing solutions to remove or alternative means of acting on prohibited content; and reporting on the content moderation actions in a transparency report.

- Though there are significant challenges that tech companies face in their counter-terrorism efforts, Tech Against Terrorism finds that tech sector responses to terrorist content are often underestimated. In its work with smaller platforms through the Tech Against Terrorism Mentorship Programme, Tech Against Terrorism has noticed that a lot of platforms go beyond what is expected or required of them. To substantiate this finding, it provides examples of the work that companies undertake beyond what is legally required, such as the rate at which they are removing terrorist or violent extremist content.

- Platforms vary greatly in their counter-terrorism policies. While some platforms choose to prohibit more general behaviour such as "illegal" actions or content, others choose to prohibit terrorism explicitly. All platforms differ in how they define and prohibit terrorism, and where they prohibit it in their content standards. Of a sample of size of 32 tech companies analysed by Tech Against Terrorism, nearly half prohibit terrorism in their content standards, and most likely located in their Community Guidelines. These platforms ranged in user base size and where they are found across the platform economy, and include platforms for communications, content storage, and social media. Of the sample size analysed, the social media platforms were the most likely to prohibit terrorism in the content standards compared to other platform types. When compared to other categories of prohibited conduct or content, such as hate speech, illegal activity, violent or graphic content, and threatening material, terrorism is prohibited least often by companies, while illegal activity is prohibited most often.

- It is important to note that 'terrorism' has many manifestations in the online sphere. These diverse manifestations are often reflected in how platforms will phrase their prohibition of terrorism in their content standards.

- For content moderation, platforms often rely on many different surfacing methods, including automated tools, human moderation teams, community or trusted flaggers, law enforcement teams, and user reporting.

- There are a range of automated tools that can be deployed during the content moderation process. Automated tools can be used at different stages of content to identify, sort, and remove content.

- Tech solutions, and especially automated solutions, are effective in scaling up otherwise time-consuming manual processes. However, to do this effectively and accurately, they need to be supported by meaningful policies and processes. Tech solutions alone cannot therefore address many underlying challenges that many smaller platforms face around building out effective and human rights-compliant moderation enforcement practices.

- Transparency reports generally demonstrate both the substance of a platform's content moderation and the nature of its cooperation with law enforcement and governments. It is important to recognise that transparency reports differ from platform to platform, and there is no one-size-fits-all solution for a comprehensive transparency report given variable capacity across the sector. Each platform will produce a report unique to its own strengths and omissions, depending on the resources available and reflecting the services it offers. This is especially true for smaller platforms with fewer resources, and who cannot be expected to produce reports including the same level of detail as larger platforms.

- Of the local platforms originating or headquartered in African Commonwealth member countries, most are likely to be marketplace or financial services platforms, while social media platforms used are international in origin. Thus, the two companies that are profiled in this report are different forms of online marketplaces. For each of the platforms, Tech Against Terrorism conducted a brief assessment of the company's policies, focusing on their prohibited conduct or content and how this is communicated to users.

- Finally, taking the collaborative industry efforts and tech sector efforts to counterterrorism into consideration, we provide our recommendations for engagement opportunities.

Recommendations

Based on Tech Against Terrorism's assessment of how Commonwealth countries can engage with existing tech sector efforts to counter terrorism, it recommends:

- Remaining up-to-date on the trends in tech sector efforts internationally, such as those analysed within this report, to inform improvements to local platforms' counter-terrorism policies and content moderation processes.

- Encouraging and supporting local platforms to sign up to available resources such as Tech Against Terrorism's mentorship programme, Knowledge Sharing Platform, and Terrorist Content Analytics Platform, all of which are outlined within this report.

- Remaining informed through following developments in the above mentioned cross-industry cooperation in the field; namely, Tech Against Terrorism, the Global Internet Forum to Counter Terrorism, and the Trust & Safety Professional Association.

Existing industry cooperation mechanisms

Although governments have passed legislation aimed at countering terrorist and other harmful online content in recent years, content moderation – including that of terrorist content – in practice remains mostly a matter for the tech sector. This entails companies drafting and applying their own rules for moderating user-generated content on their platforms in line with their values, business interests, or when they voluntarily comply with industry standards and legal enforcement. This has also historically meant that tech platforms have pioneered the strategic response to terrorist use of the internet and other harmful and illegal activity.

The predominance of self-regulation, coupled with increased public pressure to address the potentially harmful online content (in particular terrorist material), has led major tech companies to establish their own deliberative bodies to oversee their content moderation, scrutinise its impact on the freedom of speech online, and to collaborate with broader industry efforts.

We discuss below the contribution of three main boards to the provision of practical capacity building and knowledge sharing for tech companies.

The Global Internet Forum to Counter Terrorism (GIFCT)

The GIFCT was founded in 2017 by Facebook, Microsoft, Twitter, and YouTube to facilitate collaboration and knowledge sharing amongst the tech sector to tackle terrorist use of the internet. Since its founding, the GIFCT, which runs its own membership programme, has grown to encompass 21 member companies. The GIFCT is also a key partner in several public-private partnerships, including the Christchurch Call to Action.[213]

Tech Against Terrorism has been one of the GIFCT's core partners since its inception and helped to launch the coalition at its inaugural workshop in San Francisco in 2017. Tech Against Terrorism has also supported the GIFCT knowledge sharing programme by organising 13 global workshops[214] and 16 e-learning webinars. Tech Against Terrorism's Mentorship Programme is also intended to assist tech companies in meeting GIFCT's membership requirements.

Past Tech Against Terrorism and GIFCT E-learning Webinars:

Webinar Topic (date)	**Organisations represented**
Using hashing technology to counter terrorist use of the internet (June 2019)	▪ Dina Hussein, Policy EMEA Facebook. ▪ John Kerl, Facebook engineering team. ▪ Dave Ranner, Camera Forensics.
Defining terrorism and terrorist organisations on tech platforms (October 2019)	▪ Dr Erin Saltman, Policy Manager EMEA, Facebook. ▪ Dr Krisztina Huszti-Orban, Senior Legal Advisor to the UN Special Rapporteur on Counterterrorism and Human Rights. ▪ Lina Cepeda, Legal Officer & ICT Coordinator, UN CTED. ▪ Chris Meserole, Research Fellow at Brookings
Drafting Terms of Service & Community Guidelines (October 2019)	▪ Dina Hussein, Policy EMEA, Facebook. ▪ Daphne Keller, Director of Intermediary Liability, The Center for Internet & Society at Stanford Law School. ▪ Alex Feerst, former Head of Legal and Head of Trust & Safety, Medium. ▪ Sebastian Koehler, Organic Content Policy, Facebook.
Mental health and content moderation (November 2019)	▪ Dina Hussein, Policy EMEA, Facebook. ▪ Prof. Maura Conway, School of Law and Government, Dublin City University & VOX-Pol Principal Investigator. ▪ Dr Zoey Reeve, Lecturer, Research Methods, Newcastle University & VOX-Pol Research Fellow.

[213] More information on the Christchurch Call can be found in Dossier C.

[214] These workshops took place in: Sydney, Australia; Brussels, Belgium; Paris, France; Berlin, Germany; Jakarta, Indonesia; Tel Aviv, Israel; Amman, Jordan; Abu Dhabi, United Arab Emirates; California, USA (x2); New York, USA; Delhi, India; London, United Kingdom (x2)

	▪ Gustavo Basualdo, Senior Program Manager, Online Safety, Microsoft.
OSINT introduction to the current Islamist terrorist landscape (March 2020)	▪ Experts from Tech Against Terrorism
Transparency reporting for smaller tech platforms (April 2020)	▪ Dr Erin Saltman, Policy Manager EMEA at Facebook. ▪ Jessica Ashooh, Director of Policy at Reddit. ▪ Emma Llanso, Director of the Free Expression Project at the Center for Democracy & Technology.
Tech sector and law enforcement engagement in countering Terrorist Use of the Internet (May 2020)	▪ Courtney Gregoire, Chief Safety Digital Officer at Microsoft. ▪ Experts from EU IRU, Europol. ▪ Jessica Marasa, Law Enforcement Response Manager at Twitch. ▪ Stephanie McCourt, Trust & Safety Outreach Lead at Facebook.
Tech Against Terrorism Mentorship Programme and Support for Smaller Platforms (October 2020)	▪ Adam Hadley, Director, Tech Against Terrorism. ▪ Nicholas Rasmussen, Executive Director, GIFCT. ▪ Johannah Lowin, Chief of Staff, GIFCT.
Content Moderation: Alternatives to Content Removal (December 2020)	▪ Johannah Lowin, Chief of Staff, Global Internet Forum to Counter Terrorism. ▪ Alex Feerst, GC, Neuralink & Advisor, Trust and Safety Professional Association. ▪ Bill Ottman, CEO, Minds. ▪ Rachel Wolbers, Public Policy Manager, Facebook Oversight Board.
The Nexus Between Violent Extremism and Conspiracy Theory Networks Online (March 2021)	▪ Emily Thompson, Associate Director, Simon Wiesnthal Center ▪ Sam Jackson, Assistant Professor at the University at Albany and author of Oath Keepers: Patriotism and the Edge of Violence in a Right-Wing Anti-government Group ▪ Marc-Andre Argentino, PhD Candidate at Concordia University, Research Fellow at the ICSR ▪ Patrick James, Facebook, Dangerous Organizations Policy
Countering Terrorist Use of Emerging Technologies: Assessing Risks of Terrorist Use of End-to-End-Encryption and Related Mitigation Strategies (April 2021)	▪ Konstantinos Komaitis, Senior Director, Policy Development & Strategy, The Internet Society ▪ Gail Kent, Head of Messenger Policy, Facebook ▪ Marc Loebekken, Legal Counsel, ProtonMail
Technical approaches to countering terrorist use of the internet: URL sharing and collaborative tech sector efforts (April 2021)	▪ Adam Hadley, Director, Tech Against Terrorism ▪ Anne Craanen, Research Analyst, Tech Against Terrorism, ▪ Audrey Alexander, Researcher and Instructor, West Point's Combating Terrorism Center
The Nuts and Bolts of Counter Narratives: What works and why? (May 2021)	▪ Sara Zeiger, Program Manager Research and Analysis, Hedayah ▪ Munir Zamir, PhD Candidate at the University of South Wales

	▪ Tarek Elgawhary, Founder, Making Sense of Islam; CEO, Coexist Research International ▪ Ross Frenett, Founder & CEO, Moonshot CVE
APAC in Focus: Regional Responses to Terrorist and Violent Extremist Activity Online (June 2021)	▪ Maya Mirchandani, Senior Fellow at the Observer Research Foundat ion, and Assistant Professor of Broadcast, Journalism and Media Studies at Ashoka University ▪ Shashi Jayakumar, Head of the Centre of Excellence for National Security at S. Rajaratnam School of International Studies ▪ Nawab Osman, Head of Counterterrorism and Dangerous Organizations, APAC at Facebook
Supporting Platforms' Content Moderation and Transparency Efforts: Existing Resources and Tools (July 2021)	▪ Erin Saltman, Director of Programming, GIFCT ▪ Fabienne Tarrant, Research Analyst and Lead on the Knowledge Sharing Platform, ▪ Arthur Bradley, OSINT Analyst, Tech Against Terrorism
United Nations' Efforts in Counterterrorism and CVE: Resolutions, Mandates and Partnerships (August 2021)	▪ Mattias Sundholm, Strategic Communications Officer and Counter-Narratives, UNCTED ▪ Akvile Giniotiene, Head of the Cybersecurity and New Technologies Unit, UNCCT ▪ Heesu Chung, Programme Analyst, Prevention of Violent Extremism, UNDP

In 2019 the GIFCT announced that it would become an organisation independent from its parent companies. This was formalised in 2020 with the hiring of its first Executive Director, Nicholas Rasmussen. The foundational goals of the new organisation include empowering the tech sector to respond to terrorist exploitation, enabling "multi-stakeholder engagement around terrorist and violent extremist misuse of the Internet", promoting dialogue with civil society, and advancing understanding of the terrorist and violent extremist landscape "including the intersection of online and offline activities."[215]

The independent GIFCT's structure is complemented by an Independent Advisory Council (IAC) made up of 21 members representing public (including intergovernmental organisations) and civil society sectors. The IAC comprises a broad range of expertise related to the GIFCT's areas of work, such as counterterrorism, digital rights, and human rights; it is chaired by a non-governmental representative, a role currently held by Bjorn Ihler, a counter-radicalisation expert and founder of the Khalifa-Ihler Institute. The four founding companies are also represented via the Operating Board, which appoints the Executive Director and provides the GIFCT's operational budget. Other members of the board include one other member company on a rotating basis, a rotating chair from the IAC, and new members that meet "leadership criteria".[216]

The GIFCT also runs the Hash-Sharing Consortium to help member companies to moderate terrorist content on their platforms. The consortium is a database of hashed[217] terrorist content. Members can add hashes of content they have previously identified as terrorist material to the database, and all participating companies are able to automatically detect terrorist material on their platforms and prevent further uploads.

[215] https://gifct.org/

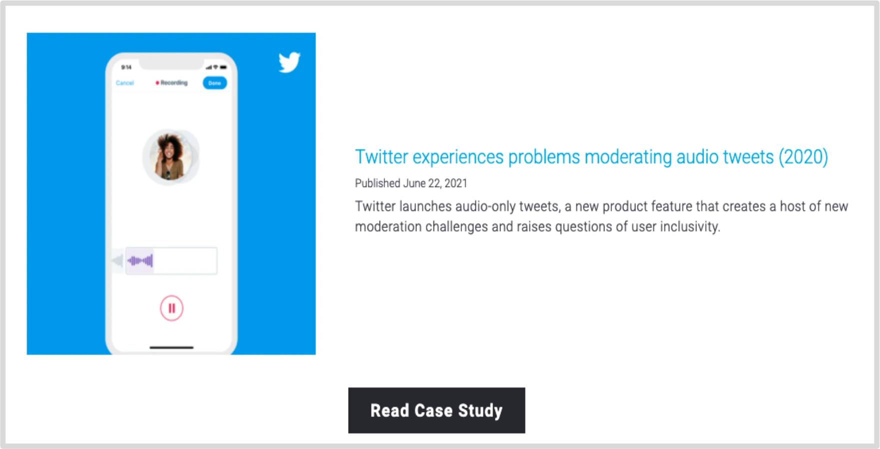
[216] https://gifct.org/

[217] "hashed" content refers to the "hashes" which are unique digital "fingerprints" of known violent terrorist imagery or terrorist recruitment videos that had been removed from their services. Read more on this here: https://gifct.org/?faqs=what-is-the-hash-sharing-consortium-and-how-does-it-work

Trust & Safety Professional Association (TSPA) and the Trust & Safety Foundation (TSF)

The Trust & Safety Professional Association (TSPA) is a non-profit, membership-based organisation[218] established to support the global community of professionals instituting the principles and policies concerned with online standards. The TSPA is a forum for professionals to connect with a network of peers, find resources for career development, and exchange best practices for navigating challenges unique to the profession. Professionals participating in this community include those who are involved in content review, policy enforcement, safety incident management, product policy development, T&S tool building, T&S analytics, and legal compliance.[219] TSPA's sister organisation, the Trust & Safety Foundation (TSF), focuses on improving society's understanding of trust and safety, including the operational practices used in content moderation, through educational programs and multidisciplinary research.[220]

The Trust & Safety Foundation hosts a library of case studies from The Copia Institute. The case studies are chosen to demonstrate how difficult it is to make online trust and safety decisions. Each case study covers a real-life dilemma, analyses the questions raised, and investigates the policy implications.[221] Please see the example provided below:



Twitter experiences problems moderating audio tweets (2020)

Published June 22, 2021

Twitter launches audio-only tweets, a new product feature that creates a host of new moderation challenges and raises questions of user inclusivity.

[218] *https://www.techdirt.com/articles/20200617/10233144735/trust-safety-professional-association-launches-this-is-important.shtml*

[219] https://www.tspa.info/about-tspa/

[220] https://www.techdirt.com/articles/20200617/10233144735/trust-safety-professional-association-launches-this-is-important.shtml

[221] https://www.tsf.foundation/case-studies

Tech Sector Efforts

Tech sector efforts to counter terrorism encompass a range of methods, including counter-terrorism policies and robust content standards, the identification of content requiring moderation as well as relevant solutions, and transparency reporting.

Companies successful in preventing terrorist or violent extremist exploitation of their platform do so by multiple means which require a range of resources such as people, policies, processes, and systems. It is important to note that smaller platforms typically have limited capacity, resources, and capability to devise the means relevant to their content moderation efforts. However, companies do work carefully to ensure that they can take the following steps to hindering terrorist and violent exploitation of their services:

- Prohibit such use in their Terms of Service or Community Guidelines, and other relevant content standards.
- Identify and detecting prohibited content, such as through user reporting or proactive monitoring, and providing solutions to remove or, alternatively, action prohibited content.
- Report on the content moderation actions in a transparency report.

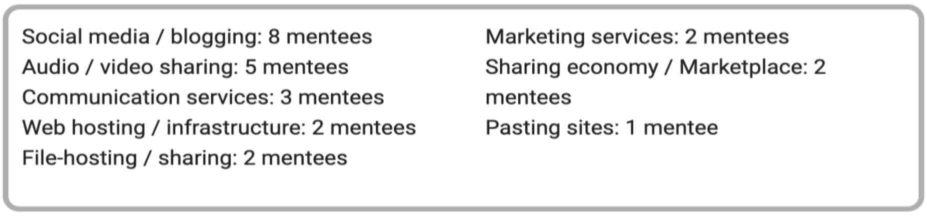
In this section, we analyse the overarching trends in tech sector efforts globally and assess the policies and practices of local platforms.

Insights from Tech Against Terrorism Mentorship

As part of Tech Against Terrorism's practical support to tech companies, it supports tech companies to grasp the threat to their platforms and makes bespoke recommendations on how to best counter it, including via its Mentorship Programme[222] and Knowledge Sharing Platform[223]. This has given us insights into how many platforms conduct their counter-terrorism policies, content moderation, and transparency reports.

Since 2018, Tech Against Terrorism has mentored 25 tech platforms to help them tackle terrorist use of their platforms without compromising human rights and the freedom of speech. Tech Against Terrorism's Mentorship Programme also supports tech platforms in strengthening transparency and accountability mechanisms around their content moderation. These platforms are representatives of the broader tech ecosystem: from small platforms run by a single person to larger tech platforms. They also represent a diverse range of online services, from social-media and video-sharing services to "sharing-economy" platforms, and each have their own content moderation approach. The mentees all face different threats in terms of terrorist and violent extremist groups attempting to exploit their services. All have demonstrated their willingness to counter terrorist and violent extremist use of their services whilst increasing transparency and accountability to their users.

Tech Against Terrorism Mentorship – Representing the broader tech sector:



Social media / blogging: 8 mentees
Audio / video sharing: 5 mentees
Communication services: 3 mentees
Web hosting / infrastructure: 2 mentees
File-hosting / sharing: 2 mentees
Marketing services: 2 mentees
Sharing economy / Marketplace: 2 mentees
Pasting sites: 1 mentee

[222] https://www.techagainstterrorism.org/2021/05/18/the-tech-against-terrorism-mentorship-2018-2021/
[223] The Knowledge Sharing Platform (KSP) is a collection of interactive tools and resources designed to support the operational needs of smaller tech platforms. It is a "one-stop shop" for companies to access practical resources to support their counterterrorism and transparency efforts. Our resources include research and analysis on terrorist use of the internet, such as the threat landscape and proscribed organisations, on global online regulation, as well as guidelines and recommendations on content standards and transparency reporting.

Through working closely with platforms, Tech Against Terrorism has seen trends in challenges with counter-terrorism policies, content moderation, as well as enforcement techniques. These trends, which are in no way exhaustive, can be summarised as follows:

Trends in challenges for counter-terrorism policies and content moderation efforts: Through its Mentorship Programme, Tech Against Terrorism has found that companies are likely to struggle with:

- Identifying and implementing explicit, operable prohibitions of terrorism and violent extremism in their Content Standards
- Correctly identifying terrorist content, in particular when terrorist actors "sanitise" their material to avoid detection
- Language capability, such as the ability to operate content moderation in multiple languages
- Responding to hostile shifts in behaviour to evade content moderation avoidance strategies
- Developing transparency reports

Trends in solutions for counter-terrorism policies and content moderation efforts: Tech sector responses to terrorist content are often underestimated. In Tech Against Terrorism's work with smaller platforms, it has noticed that a lot of platforms go beyond what is expected or required of them. For example, they will often go beyond prohibiting what is considered illegal content in their jurisdictions, such as by providing more detailed prohibitions. In addition, they have in place elaborate user reporting mechanisms or appeal procedures as well as sophisticated content moderation solutions, which many account for in their transparency reports. These techniques and the trends of how they are accounted for are outlined in the section below.

In addition to the policies and reporting mechanisms in place, platforms remove reported content at a remarkable rate. According to the Gap Analysis on Technical Approaches to Counter Terrorist of the Internet,[224] which Tech Against Terrorism published in July in partnership with the GIFCT, most large platforms automatically remove 95%+ of terrorist content and most smaller platforms respond to takedown requests within hours of being notified. Data from the Terrorist Content Analytics Platform[225] shows that since November 2020 96% of URLs pointing to verified terrorist content was removed by smaller platforms.

State of play: overarching trends in counter-terrorism policies

Below, we analyse tech companies' practical commitments to counter-terrorism policy, content moderation, and transparency reporting.

Counter-terrorism Policy Benchmarks

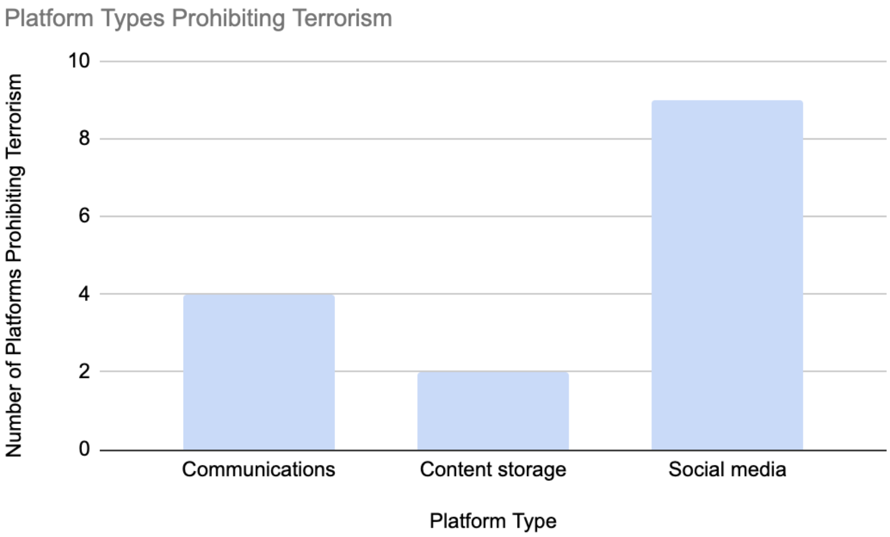
Platforms vary greatly in their counter-terrorism policies. While some platforms choose to prohibit more general behaviour such as "illegal" actions or content, others choose to explicitly prohibit terrorism. All platforms differ in how they define and prohibit terrorism, and where in their content standards they prohibit it.

Of a sample of size of 32 tech companies that Tech Against Terrorism analysed, 15 prohibit terrorism in their content standards. These platforms ranged in userbase size and location across the platform economy, and comprise platforms for communications, content storage, and social media. Below we outline how these platforms prohibit banned categories of content or conduct in their content standards.

[224] https://gifct.org/wp-content/uploads/2021/07/GIFCT-TAWG-2021.pdf?utm_source=Tech+Against+Terrorism&utm_campaign=dad15fffda-EMAIL_CAMPAIGN_2019_03_24_07_51_COPY_01&utm_medium=email&utm_term=0_cb464fdb7d-dad15fffda-162810151

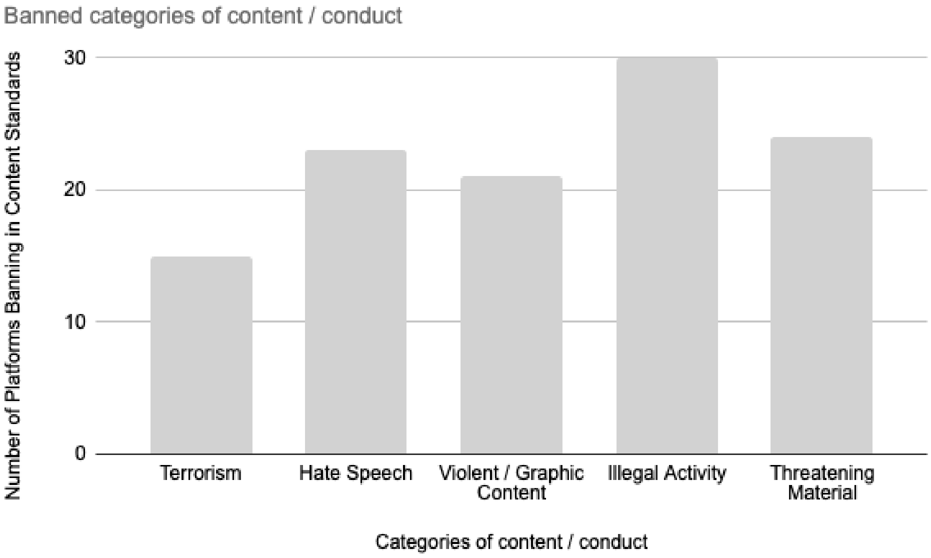
[225] https://www.terrorismanalytics.org/

Of the 32 platforms analysed, the social media platforms were the most likely to prohibit terrorism in the content standards compared to other platform types. Communications platforms[226] and content storage platforms did so less often. This is demonstrated in the chart below:



While not all tech companies prohibit terrorism, they often prohibit other types of categories of content or conduct. Below we have highlighted the number of platforms banning categories of content or conduct which includes terrorism, hate speech, violent or graphic content, illegal activity, and threatening material. Based on this analysis, most platforms had chosen to prohibit illegal activity. However, it is important to note that most of the tech companies prohibit more than one of the named banned categories of content or conduct.

Based on the figure below, most tech companies prohibited illegal activity, while terrorism was prohibited the least often. Though illegal activity generally can act as an umbrella term for terrorism or other conduct, it is important for platforms to clearly prohibit the explicit conduct that may fall under such category, particularly terrorism. In the section below, we portray in what ways platforms are prohibiting terrorism in their content standards.



[226] This entails messenger platforms such as WhatsApp, Line and Signal.

Prohibiting terrorism

An explicit prohibition of terrorism in content standards demonstrates the seriousness with which a platform confronts the threat of terrorist use. It is important to note that ‘terrorism’ has many manifestations in the online sphere. These include, but are not limited to, using a platform to praise or glorify terrorist entities and activities, promoting terrorists’ agenda through sharing propaganda or symbols, and providing logistical or material support to terrorist entities such as through recruitment or funding campaigns.

We identify below some of these uses and the corresponding extracts from select companies’ content standards to provide insight into how companies prohibit

a) Use of the platform by terrorist organisations and/or members:

b) Support for terrorist activities and/or those engaged in terrorism (material or otherwise)

c) Praise of terrorist activities and/or those engaged in terrorism:

d) Promotion of terrorist activities and/or those engaged in terrorism (promotion includes posting content produced by terrorist entities)

e) Recruitment for terrorist organisations and/or terrorist activities

f) Engaging in or threatening or inciting acts of terrorism on behalf of a terrorist organization on the platform

Prohibiting the use of the platform by terrorist organisations and/or members

Facebook[227]	“In an effort to prevent and disrupt real-world harm, we do not allow any organisations or individuals that proclaim a violent mission or are engaged in violence to have a presence on Facebook. This includes organisations or individuals involved in the following:” ▪ Terrorist activity ▪ Organised hate ▪ Mass murder (including attempts) or multiple murder ▪ Human trafficking ▪ Organised violence or criminal activity
Twitter[228]	“There is no place on Twitter for violent organizations, including terrorist organizations, violent extremist groups, or individuals who affiliate with and promote their illicit activities.”
YouTube[229]	“These [violent criminal] organizations are not allowed to use YouTube for any purpose, including recruitment.”
Snapchat[230]	“Terrorist organisations are prohibited from using our platform, and we have no tolerance for content that advocates or advances terrorism.”
Google Drive[231]	“Terrorist organizations are not permitted to use this product for any purpose, including recruitment.”

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

[228] https://help.twitter.com/en/rules-and-policies/violent-groups

[229] https://support.google.com/youtube/answer/9229472?hl=en&ref_topic=9282436

[230]https://www.google.com/url?q=https://snap.com/en-GB/community-guidelines&sa=D&source=editors&ust=1629812895838000&usg=AOvVaw3q5LAHsti7OYZ28I4kBIOw

[231] https://support.google.com/docs/answer/148505#zippy=%2Cdangerous-and-illegal-activities%2Charassment-bullying-and-threats%2Caccount-hijacking%2Caccount-inactivity%2Ccircumvention%2Cchild-sexual-abuse-and-exploitation%2Chate-speech%2Cmisleading-content%2Cmalware-and-similar-malicious-content%2Cimpersonation-and-misrepresentation%2Cregulated-goods-and-services%2Cphishing%2Cpersonal-and-confidential-information%2Cnon-consensual-explicit-imagery-ncei%2Cunauthorized-images-of-minors%2Cterrorist-activities%2Csystem-interference-and-abuse%2Cspam%2Csexually-explicit-material%2Cviolence-and-gore%2Ccontent-use-and-submission

Vimeo[232]	“Certain users may not use our services, regardless of their content. These are: gangs, hate groups, terror organizations, members of the foregoing, and persons who are subject to U.S. sanctions. In addition, if you are in a country that is subject to comprehensive U.S. sanctions, you may not purchase software services or hardware from us.”

Prohibiting support for terrorist activities and/or those engaged in terrorism :

Facebook[233]	“We also remove content that expresses support or praise for groups, leaders or individuals involved in these activities [including terrorist activity].” “We do not allow coordination of support for any of the above organisations [including terrorist organizations] or individuals or any acts committed by them.” “We do not allow content that praises, supports or represents events that Facebook designates as terrorist attacks, hate events, mass murders or attempted mass murders, serial murders, hate crimes and violating events.”
Twitter[234]	“Examples of the types of content that violate this policy include, but are not limited to] providing or distributing services (e.g., financial, media/propaganda) to further a violent organization’s stated goals”; and “using the insignia or symbol of violent organizations to promote them or indicate affiliation or support.”
YouTube[235]	“Content intended to praise, promote, or aid violent criminal organizations is not allowed on YouTube.”
Tumblr[236]	“We don’t tolerate content that promotes, encourages, or incites acts of terrorism. That includes content which supports or celebrates terrorist organizations, their leaders, or associated violent activities.”
Vimeo[237]	“You may not upload content that: Promotes or supports terror or hate groups.” “We do not allow content from hate or terror groups that aims to spread propaganda designed to radicalise and recruit people or aid and abet attacks”.

Prohibiting the praise of terrorist activities and/or those engaged in terrorism:

Facebook[238]	“We also remove content that expresses support or praise for groups, leaders or individuals involved in these activities [including terrorist activity].” “We do not allow content that praises any of the above organisations or individuals or any acts committed by them.” “We do not allow content that praises, supports or represents events that Facebook designates as terrorist attacks, hate events, mass murders or attempted mass murders, serial murders, hate crimes and violating events.”
YouTube[239]	“Content intended to praise, promote, or aid violent criminal organizations is not allowed on YouTube.”
Tumblr[240]	“We don’t tolerate content that promotes, encourages, or incites acts of terrorism. That includes content which supports or celebrates terrorist organizations, their leaders, or associated violent activities.”
Google Drive[241]	“We’ll also take action against the user for content related to terrorism, such as promoting terrorist acts, inciting violence, or celebrating terrorist attacks.”

[232] https://vimeo.com/help/guidelines#restricted_users
[233] https://www.facebook.com/communitystandards/dangerous_individuals_organizations
[234] https://help.twitter.com/en/rules-and-policies/violent-groups
[235] https://support.google.com/youtube/answer/9229472?hl=en&ref_topic=9282436
[236] https://www.tumblr.com/policy/en/community
[237] https://vimeo.com/help/guidelines
[238] https://www.facebook.com/communitystandards/dangerous_individuals_organizations
[239] https://support.google.com/youtube/answer/9229472?hl=en&ref_topic=9282436
[240] https://www.tumblr.com/policy/en/community
[241] https://support.google.com/docs/answer/148505#zippy=%2Cdangerous-and-illegal-activities%2Charassment-bullying-and-threats%2Caccount-hijacking%2Caccount-inactivity%2Ccircumvention%2Cchild-sexual-abuse-and-exploitation%2Chate-speech%2Cmisleading-

Prohibiting the promotion of terrorist activities and/or those engaged in terrorism (promotion includes posting content produced by terrorist entities):

Facebook[242]	“We do not allow symbols that represent any of the above organisations or individuals to be shared on our platform without context that condemns or neutrally discusses the content.”
Twitter[243]	“Under this policy, you can’t affiliate with and promote the illicit activities of a terrorist organization or violent extremist group.” “[Violating our policy includes] using the insignia or symbol of violent organizations to promote them or indicate affiliation or support”. “[Violating our policy includes] “engaging in or promoting acts on behalf of a violent organization”.
YouTube[244]	“Don’t post [...] Content that depicted the signia, logos, or symbols of violent criminal or terrorist organizations in order to praise or promote them”. “Don’t post [...] Content produced by violent criminal or terrorist organizations”.
Tumblr[245]	“We don’t tolerate content that promotes, encourages, or incites acts of terrorism.”
Dropbox[246]	“You must not publish or share materials [...] that contain extreme acts of violence or terrorist activity, including terror propaganda.”
Google Drive[247]	“We’ll also take action against the user for content related to terrorism, such as promoting terrorist acts, inciting violence, or celebrating terrorist attacks.”

Prohibiting **recruitment** for terrorist organisations and/or terrorist activities:

Twitter[248]	“[Violating our policy includes] “recruiting for a violent organization”
YouTube[249]	“Content intended to praise, promote, or aid violent criminal organizations is not allowed on YouTube. These organizations are not allowed to use YouTube for any purpose, including recruitment.” “Don’t post […] Content aimed at recruiting new members to violent criminal or terrorist organizations.” “[Some examples of content that’s not allowed on YouTube] … Content directing users to sites that espouse terrorist ideology, are used to disseminate prohibited content, or are used for recruitment.”

content%2Cmalware-and-similar-malicious-content%2Cimpersonation-and-misrepresentation%2Cregulated-goods-and-services%2Cphishing%2Cpersonal-and-confidential-information%2Cnon-consensual-explicit-imagery-ncei%2Cunauthorized-images-of-minors%2Cterrorist-activities%2Csystem-interference-and-abuse%2Cspam%2Csexually-explicit-material%2Cviolence-and-gore%2Ccontent-use-and-submission

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

[243] https://help.twitter.com/en/rules-and-policies/violent-groups

[244] https://support.google.com/youtube/answer/9229472?hl=en&ref_topic=9282436

[245] https://www.tumblr.com/policy/en/community

[246] https://www.dropbox.com/acceptable_use

[247] https://support.google.com/docs/answer/148505#zippy=%2Cdangerous-and-illegal-activities%2Charassment-bullying-and-threats%2Caccount-hijacking%2Caccount-inactivity%2Ccircumvention%2Cchild-sexual-abuse-and-exploitation%2Chate-speech%2Cmisleading-content%2Cmalware-and-similar-malicious-content%2Cimpersonation-and-misrepresentation%2Cregulated-goods-and-services%2Cphishing%2Cpersonal-and-confidential-information%2Cnon-consensual-explicit-imagery-ncei%2Cunauthorized-images-of-minors%2Cterrorist-activities%2Csystem-interference-and-abuse%2Cspam%2Csexually-explicit-material%2Cviolence-and-gore%2Ccontent-use-and-submission

[248] https://help.twitter.com/en/rules-and-policies/violent-groups

[249] https://support.google.com/youtube/answer/9229472?hl=en&ref_topic=9282436

Google Drive[250]	“Terrorist organizations are not permitted to use this product for any purpose, including recruitment.”

Prohibiting engaging in or threatening or inciting acts of terrorism on behalf of a terrorist organization on the platform:

Twitter[251]	“[Violating our policy includes] engaging in or promoting acts on behalf of a violent organization;” “You may not threaten or promote terrorism or violent extremism.”
YouTube[252]	“Don’t post […] Content depicting hostages or posted with the intent to solicit, threaten, or intimidate on behalf of a violent criminal or terrorist organization”
Tumblr[253]	“We don’t tolerate content that […] incites acts of terrorism.”
Snapchat[254]	“Terrorist organisations are prohibited from using our platform, and we have no tolerance for content that advocates or advances terrorism.”
Google Drive[255]	“We’ll also take action against the user for content related to terrorism, such as promoting terrorist acts, inciting violence, or celebrating terrorist attacks”

Where terrorism is prohibited

Platforms will prohibit terrorism and other illegal use or content in different areas of their content standards. Of the 32 tech companies analysed by Tech Against Terrorism analysed,[256] 15 platforms prohibit terrorism in the content standards, of which most prohibited terrorism in their Community Guidelines. Two platforms prohibited terrorism in both their Community Guidelines and Terms of Service.

[250] https://support.google.com/docs/answer/148505#zippy=%2Cdangerous-and-illegal-activities%2Charassment-bullying-and-threats%2Caccount-hijacking%2Caccount-inactivity%2Ccircumvention%2Cchild-sexual-abuse-and-exploitation%2Chate-speech%2Cmisleading-content%2Cmalware-and-similar-malicious-content%2Cimpersonation-and-misrepresentation%2Cregulated-goods-and-services%2Cphishing%2Cpersonal-and-confidential-information%2Cnon-consensual-explicit-imagery-ncei%2Cunauthorized-images-of-minors%2Cterrorist-activities%2Csystem-interference-and-abuse%2Cspam%2Csexually-explicit-material%2Cviolence-and-gore%2Ccontent-use-and-submission

[251] https://help.twitter.com/en/rules-and-policies/violent-groups

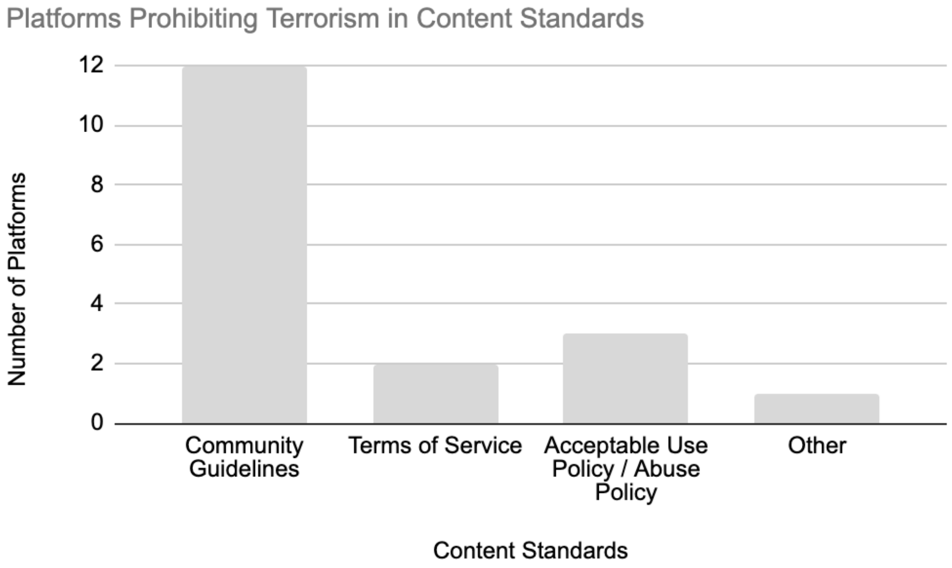
[252] https://support.google.com/youtube/answer/9229472?hl=en&ref_topic=9282436

[253] https://www.tumblr.com/policy/en/community

[254] https://snap.com/en-GB/community-guidelines

[255] https://support.google.com/docs/answer/148505#zippy=%2Cdangerous-and-illegal-activities%2Charassment-bullying-and-threats%2Caccount-hijacking%2Caccount-inactivity%2Ccircumvention%2Cchild-sexual-abuse-and-exploitation%2Chate-speech%2Cmisleading-content%2Cmalware-and-similar-malicious-content%2Cimpersonation-and-misrepresentation%2Cregulated-goods-and-services%2Cphishing%2Cpersonal-and-confidential-information%2Cnon-consensual-explicit-imagery-ncei%2Cunauthorized-images-of-minors%2Cterrorist-activities%2Csystem-interference-and-abuse%2Cspam%2Csexually-explicit-material%2Cviolence-and-gore%2Ccontent-use-and-submission

[256] The sample of tech platforms range in size and are found across the platform economy, including communications, content storage, and social media.



Glossary of Content Standards

The below table outlines key definitions for content standards such as Terms of Service, Community Guidelines, and Acceptable Use Policy.

Terms of Service	A legally binding agreement,[257] required of all platforms that store personal data for a user, containing rules that must be observed in order to use a service.
Community Guidelines	A set of foundational principles aiming to balance self-expression with safety, to protect both users and the platform (usually found on social media platforms).
Acceptable Use Policy	A set of foundational principles aiming to balance self-expression with safety, to protect both users and the platform (usually found on social media platforms).

Platforms often use some of the terminologies above interchangeably. In particular, the Acceptable Use Policy and Community Guidelines often overlap as they similarly delimit proper use of the service. In addition, both of these sections can often be found within a platform's Terms of Service under one combined 'legal' page. There is a clear trend of counter-terrorism policies being located in the Community Guidelines rather than the Terms of Service.

Content moderation

The below section discusses content moderation and focuses on the different types and stages of content moderation, commonly used techniques to identify terrorist content, typical solutions to moderating content beyond simple removal , and cross-industry mechanisms that provide practical support, such as those provided by Tech Against Terrorism and the GIFCT,

[257] The terms of service are the legal agreements between a service provider and a person who wants to use that service. They are mainly a contractual agreement, as only a few legislations mandate them. This is beginning change, however. Where they are not yet mandated by legislation, the terms of service act as a way for platforms to protect themselves from legal liability.

Content moderation consists both in the enforcement of a platform's own guidelines on what is and is not acceptable, as well as in compliance with requests made by government agencies and law enforcement. Content moderation is often assumed to be synonymous with content removal, whereas there are in fact many different editorial processes that are employed by platforms in enforcing their guidelines.

Detecting Content

Detection is the process of monitoring and identifying potentially problematic user-generated content. Different moderation methods are relevant at different phases of publication, and thus the status of the content (published, un-published, in the process of being published and so on) determines the appropriate action. Legal scholar and content moderation expert Kate Klonick (2017) has distinguished between three chronological stages:[258]

1. *Ex Ante Content Moderation:* Meaning before the fact, or before the content is uploaded. It takes place between commencement and completion of uploading.

2. *Ex Post Proactive Manual Content Moderation:* This process refers to platforms that proactively seek and remove content published previously but undetected by the platform's moderation systems. Importantly, Facebook has employed this method to identify previously posted terrorist content still present on the platform. The majority of this content was posted prior to Facebook's deployment of automated solutions (including image hashing) to identify terrorist content.

3. *Ex Post Reactive Manual Content Moderation:* This refers to published content subjected to review by human moderators, often relatively shortly after being identified or flagged by moderators, users, or automated detection solutions.

Platforms often rely on many different identification methods, including automated tools, human moderation teams, community or trusted flaggers, law enforcement teams, and user reporting.

Detection Strategies

While some platforms use user reporting to detect content, others also utilise more proactive monitoring and automated tools.

User reporting

As part of the *Ex Post Reactive Manual Content Moderation*, many platforms depend on their user reporting mechanisms.

Below is a diagram outlining the user journey when reporting terrorist content on Facebook, in an example of reporting tools for terrorist content - such examples highlight how platforms structure their evolving user reporting features. It is important that all reporting categories refer clearly to a platform's content standards and that users are aware of what platforms mean by each reporting category to ensure the most accurate possible reporting by and evaluation by content moderation teams.

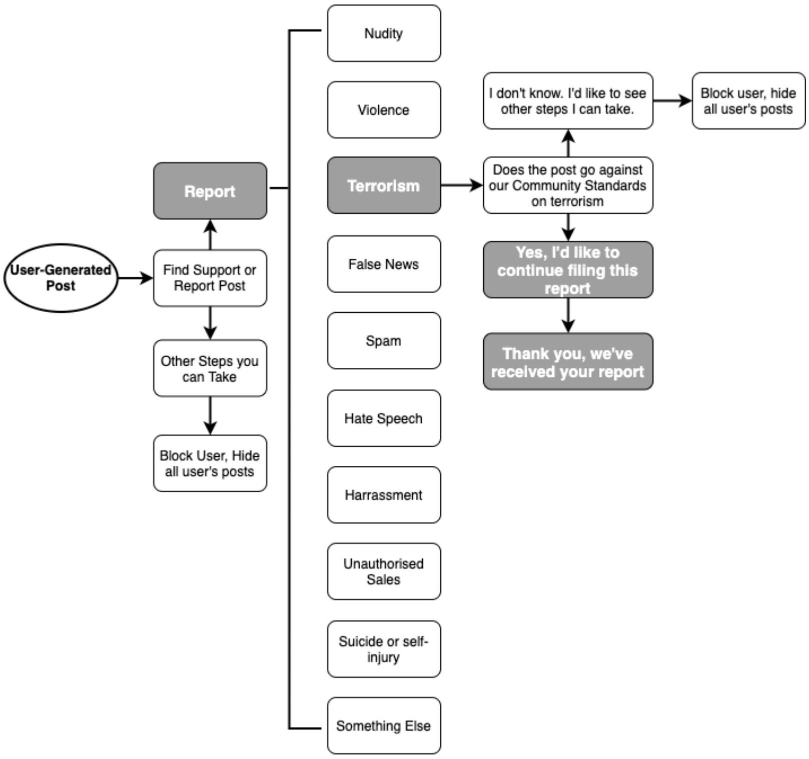
Designing and implementing user reporting requires multi-perspectival thought to be given to policy, processes, and systems, of which smaller platforms are not necessarily capable. Tech Against Terrorism's observations of user reporting and recommendations for best practice are outlined as follows:

- Have a user reporting mechanism to allow for users to report content they believe to be in violation of the Content Standards. The following options can be used for user reporting:
    - Direct flagging, via a report button available next to each content or user profile.
    - Reporting form.
    - Email reporting via dedicated email address.
    - In the case of email reporting or a reporting form, platforms should request the following information: date of upload and report, the URL address and a description of the alleged prohibited content, as well as the username of the uploader.

[258] THE NEW GOVERNORS: THE PEOPLE, RULES, AND PROCESSES GOVERNING ONLINE SPEECH, Kate Klonick, 1636-1638.

- Ensure that terrorist content is a category on its own for user reporting:
  - A dedicated reporting category helps prioritising and segmenting reviews of user reports by the Trust & Safety team, ensuring that terrorist content is dealt with swiftly.
  - This will also facilitate collecting data for transparency reports.
  - In general, it demonstrates that such content is considered a serious and dangerous category by the platform.

Reporting terrorist content on Facebook [259]


User-Generated Post
Find Support or Report Post
Report
Other Steps you can Take
Block User, Hide all user's posts
Nudity
Violence
Terrorism
False News
Spam
Hate Speech
Harrassment
Unauthorised Sales
Suicide or self-injury
Something Else
Does the post go against our Community Standards on terrorism
I don't know. I'd like to see other steps I can take.
Block user, hide all user's posts
Yes, I'd like to continue filing this report
Thank you, we've received your report

Automated tools

There are a range of automated tools that can be deployed during the content moderation process.[260] Automated tools can be used at different stages of content, to identify, sort, and remove content.

According to Tech Against Terrorism's recently published Gap Analysis report,[261] tech solutions, and especially automated solutions, are effective in scaling up otherwise time-consuming manual processes. However, to do this effectively and accurately, they need to be based on substantive policies and processes. Tech solutions alone cannot therefore address the many underlying challenges that many smaller platforms face in building out effective and human rights-compliant practices of moderation enforcement.[262]

[259] For more user reporting path diagrams of other tech companies, please see our Knowledge Sharing Platform.

[260]https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/how-automated-tools-are-used-in-the-content-moderation-process

[261] https://gifct.org/wp-content/uploads/2021/07/GIFCT-TAWG-2021.pdf

[262] See Alexander Stamos, "Prepared written testimony and statement for the record before the US House of Representatives Committee on Homeland Security, Subcommittee on Intelligence and Counter-terrorism on 'Artificial Intelligence and Counter-terrorism: Possibilities and Limitations'," June 25, 2019, https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/stamos_written_testimony_-_house_homeland_security_committee_-_ai_and_counter-terrorism.pdf

Below we outline some of the most widely used automated tools and methods in content moderation as identified by New America's Open Technology Institute's analysis in *Everything in Moderation.*[263] The tools below include digital hash technology and image recognition.

Digital hashing technology

In digital hash technology, images and videos are converted from an existing database into a grayscale format. They are then overlaid onto a grid and each square is assigned a numerical value. The designation of a numerical value converts the square into a hash, or digital signature, which remains tied to the image or video and can be used to identify further appearances of the content either during ex-ante moderation or ex-post proactive moderation.[264]

Image recognition

During ex-post proactive moderation, image recognition tools can identify specific objects within an image, such as weaponry, and decide based on factors including user experience and risk whether the image should be flagged to a human for review. According to New America's Open Technology Institute, automated image recognition tools are currently employed by several internet platforms, as they help filter through and prioritize cases for human moderators.[265]

Limitations to Detections Strategies

It is important to recognise that automated tools used for content moderation are limited in a number of ways. There are several well-reported risks associated with using automated data-driven solutions to counter terrorist use of the internet.[266] Most of this concerns the error rates and false positives that such systems produce, largely as a result of automated solutions not being able to account for context or nuance.

As per Tech Against Terrorism's Gap Analysis Report, the main risks with such tools include:[267]

- Negative impact on freedom of speech by accidentally removing legitimate content through counter-terrorism policies. Some studies suggest that such error rates predominantly affect minority groups.[268]
- Unwarranted surveillance
- Lack of transparency and accountability in the development process, which precludes interrogation by external reviewers.
- Accidental deletion of digital evidence content, much of which is crucial in terrorism and war crime trials.

In Tech Against Terrorism's assessment, the challenges posed to human rights compliance by the automation of content moderation are predominantly the result of developers having insufficient regard for these risks, or not having access to accurate training data or guidance from subject matter and human rights experts. In some cases, there seems to be little coordination with external stakeholders, including subject matter experts and civil society, before such solutions are brought to the market.

Tech Against Terrorism's Gap Analysis Report also highlights that there is a gap with regards to what many stakeholders expect tech solutions to be able to do and what they are effective at doing. What

[263]https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/

[264] Klonick, "The New Governors: The People, Rules, and Processes Governing Online Speech". Quoted in: https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/how-automated-tools-are-used-in-the-content-moderation-process

[265] Accenture, Content Moderation: The Future is Bionic, 2017, source. Quoted in https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/how-automated-tools-are-used-in-the-content-moderation-process

[266] Dia Kayyali, "Vital Human Rights Evidence in Syria is Disappearing from YouTube," WITNESS, August 2017, link; Joint Report of Electronic Frontier Foundation, Witness, and Syrian Archive, "Caught in the Net: The Impact of 'Extremist' Speech Regulations on Human Rights Content," May 30, 2019, https://blog.witness.org/2017/08/vital-human-rights-evidence-syria-disappearing-youtube/.

[267] https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/the-limitations-of-automated-tools-in-content-moderation

[268] https://www.article19.org/resources/the-global-impact-of-content-moderation/

is needed is clarification of what role automated solutions can and should play based on current capabilities. In so doing, we should move away from the flawed notion that "artificial intelligence" can effectively tackle online terrorist content without adequate human guidance. Instead, we should build consensus around exactly which tasks and workstreams technology should support, with reference to both the identification and moderation of content.[269]

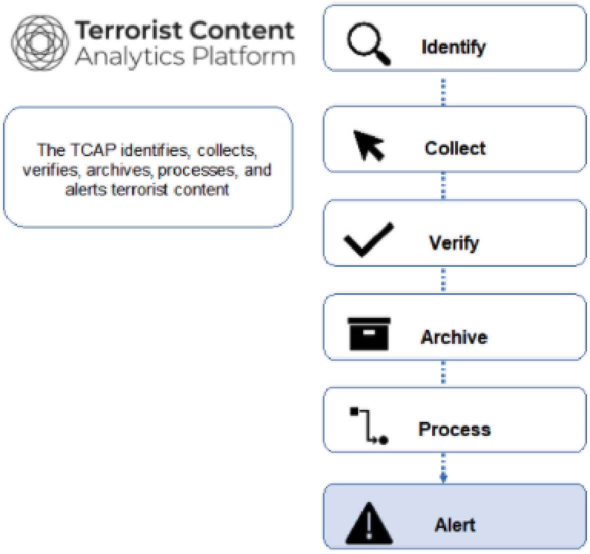
For a more in depth understanding of automated content moderation tools and their impact on human rights and freedom of expression, please see Tech Against Terrorism's dossier "Countering terrorist use of the internet whilst respecting human rights".

Cross-industry mechanisms for content moderation: Tech Against Terrorism resources to support with detection

The Terrorist Content Analytics Platform: The Terrorist Content Analytics Platform (TCAP)[270] is a secure online tool that automates the detection and analysis of verified terrorist content on smaller internet platforms. The TCAP alerts platforms to branded content associated with designated (far-right and Islamist) terrorist organisations, archives the material, and facilitates discussion between platforms, civil society, and academia to improve classification and moderation of illegal content.

This dataset of verified terrorist content supports smaller platforms in making better content moderation decisions as it assists with the swift detection of content previously identified as terrorist in nature. The platform alerts tech companies to terrorist content which they then compare to their own Terms of Service to determine whether it should be actioned. It is important to note that these alerts are made on an advisory basis only.

Alerting terrorist and violent extremist content: Terrorist Content Analytics Platform


Terrorist Content Analytics Platform
The TCAP identifies, collects, verifies, archives, processes, and alerts terrorist content
Identify
Collect
Verify
Archive
Process
Alert

- *The Knowledge Sharing Platform*

The Knowledge Sharing Platform (KSP) is a collection of interactive tools and resources designed to support the operational needs of smaller tech platforms. It is a "one stop shop" for companies to

[269] https://gifct.org/wp-content/uploads/2021/07/GIFCT-TAWG-2021.pdf?
[270] terrorismanalytics.org

access practical resources to support their counterterrorism and transparency efforts. Tech Against Terrorism's resources include research and analysis concerning terrorist use of the internet, (including information on the threat landscape and proscribed organisations) global online regulation, as well as guidelines and recommendations for content standards and transparency reporting.

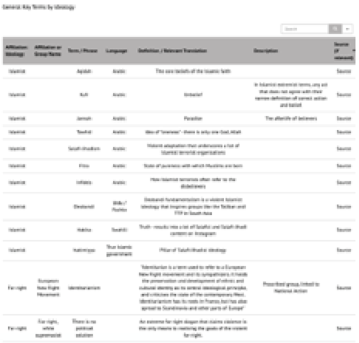
The KSP supports platforms by including key terms, logos, symbols, all of which can inform the detection of content moderation teams. Examples of this are provided below:

150+

150+ symbols and visual identifiers for 20+ designated terrorist groups for the far-right, far-left, and Islamist terrorist ideologies

900+

900+ key terms and phrases in English and other relevant translations for 20+ designated terrorist groups for the far-right, far-left, and Islamist terrorist groups



General Key Terms by Ideology

Search

Affiliation: Ideology	Affiliation or Group Name	Term / Phrase	Language	Definition / Relevant Translation	Description	Source (if relevant)
Islamist		Aqidah	Arabic	The core beliefs of the Islamic faith		Source
Islamist		Kufr	Arabic	Unbelief	In Islamist extremist terms, any act that does not agree with their narrow definition of correct action and belief	Source
Islamist		Jannah	Arabic	Paradise	The afterlife of believers	Source
Islamist		Tawhid	Arabic	Idea of "oneness" - there is only one God, Allah		Source
Islamist		Salafi-Jihadism	Arabic	Violent adaptation that underscores a lot of Islamist terrorist organisations		Source
Islamist		Fitra	Arabic	State of pureness with which Muslims are born		Source
Islamist		Infidels	Arabic	How Islamist terrorists often refer to the disbelievers		Source
Islamist		Deobandi	Urdu / Pashto	Deobandi fundamentalism is a violent Islamist ideology that inspires groups like the Taliban and TTP in South Asia		Source
Islamist		Hakika	Swahili	Truth - results into a lot of Salafist and Salafi-Jihadi content on Instagram		Source
Islamist		Hakimiyya	True Islamic government	Pillar of Salafi-Jihadist ideology		Source
Far-right	European New Right Movement	Identitarianism		"Identitarian is a term used to refer to a European New Right movement and its sympathisers. It holds the preservation and development of ethnic and cultural identity as its central ideological principle, and criticises the state of the contemporary West. Identitarianism has its roots in France, but has also spread to Scandinavia and other parts of Europe"	Proscribed group, linked to National Action	Source
Far-right	Far-right, white supremacist	There is no political solution		An extreme far-right slogan that claims violence is the only means to realising the goals of the violent far-right.		Source

- *GIFCT Hash-sharing*

The Global Internet Forum to Counter Terrorism (GIFCT) hash-sharing base is not a Tech Against Terrorism resource, however as a key partner of the GIFCT Tech Against Terrorism mentors tech companies to bring them to a certain industry standard to be eligible for GIFCT Membership (a precursor to gaining access to the GIFCT hash-sharing database).

The GIFCT's Hash-Sharing Consortium supports their member companies to moderate terrorist content on their platforms. The consortium is a database of hashed terrorist content. Members can add hashes of content they have previously identified as terrorist material to the database. All companies using it are able to automatically detect terrorist material on their platforms and prevent further uploads. [271]

Whilst the GIFCT states that "each consortium member can decide how they would like to use the database based on their own user terms of service", critics have raised concerns over the lack of transparency surrounding the use of the database and the removal of content to which it contributes .[272] However, the GIFCT has to date published two transparency reports, explaining the hash-sharing database and the type of content added to it.[273] The GIFCT said in its 2020 report that the hash-sharing database contained content in the following categories:

[271] https://gifct.org/
[272] https://www.voxpol.eu/one-database-to-rule-them-all/
[273] https://gifct.org/transparency/

- Imminent Credible Threat: 0.1%
- Graphic Violence Against Defenceless People: 16.9%
- Glorification of Terrorist Acts: 72%
- Radicalisation, Recruitment, Instruction: 2.1%
- Christchurch attack and Content Incident Protocols [274](Christchurch, 6.8% Halle attack, 2% Glendale attack 0.1%)

Academic and online regulation expert, Evelyn Douek,[275] has used the GIFCT as an example when cautioning against a phenomenon she calls "content cartels". In her analysis, Douek emphasises the perceived risk of collaborative industry arrangements involving both larger and smaller companies, where "already powerful actors" can gain further power as they are able to set content regulation standards for the smaller platforms. In particular, she argues that such arrangements leave little room for challenging the standards they set – including, in some cases, what they consider to be terrorist or harmful content.[276]

Content moderation strategies

Below we have listed some of the common strategies employed by tech companies as content moderation solutions in lieu of wholesale content removal or account deletion.

Hiding Content

Hiding content either partially or from selected users is a way of temporarily removing reported content which avoids suppressing potentially legal, legitimate, and/or allowed content across the entire platform. For example, this might be for users from a more vulnerable demographic or from a country where the content violates laws (but does not in others).

Disengagement

Disengagement schemes are a means of suppressing content/users by decreasing engagement or activity around a post/account without actually taking them off the platform. This can be done by manipulating a site's functioning to not promote such content further, or by demoting users that might otherwise be prominently identified. Disengagement can be seen as distinct from hiding content as these schemes seek to demote a piece of content or a user's prominence on a platform across the platform, rather than only for certain sections of users.

Educational or Communications Tactics

Educational or comms-based tactics seek to provide users with extra information around a piece of content so that they can decide for themselves whether they want to see it or engage with it, thereby relieving the burden of decision on the platform. Such strategies can be understood as a means of empowering users, but nonetheless requires intervention by a platform to adjudicate what content merits the provision of further information and what that further information should be

Community Empowerment Initiatives

Features focused on community empowerment rely on users to curate the online space that they want to see. These strategies are ubiquitous among community-reliant platforms or platforms that have smaller teams of content moderators. Such strategies are particularly relevant for content that is offensive but does not strictly violate a platform's content standards.

A way of deciding what content stays on a platform that belongs to the community-reliant approach is often called distributed moderation. In most cases, platforms base their moderation on legal frameworks or their own policies but in some instances abstain completely from the process of content adjudication and remit the matter wholly to users.

[274] The Content Incident Protocol (CIP) is a process by which GIFCT member companies quickly become aware of, assess, and address potential content circulating online resulting from an offline terrorist or violent extremist event. See more information here: https://gifct.org/content-incident-protocol/

[275] https://cyber.harvard.edu/people/evelyn-douek

[276] https://knightcolumbia.org/content/the-rise-of-content-cartels

Transparency reporting

Company transparency reports enable the tech industry to demonstrate to governments and the wider public how they moderate content on their platforms. Transparency reports may also illuminate whether tech companies or governments abuse content takedown mechanisms for purposes other than counterterrorism and the response to the legitimate concerns of users and civil society.

Tech Against Terrorism sees three main benefits to transparency reporting:

- Reinforcement company values while easing concerns for users' privacy
- Raising awareness of the volume and extent of government requests, thus making it easier to hold governments accountable for potential infringements,
- Contribution to the wider debate about how content can be regulated without being removed

It is important to recognise that transparency reports differ from platform to platform, and there is no one-size-fits-all solution for a comprehensive transparency report due to variable capacities. Each platform will have a unique report with its own strengths and omissions, depending on the resources available and reflecting the services it offers.[277] This is especially true for smaller platforms with fewer resources, and which cannot be expected to produce reports including the same level of detail as larger platforms.

In an attempt to highlight some of the commonly used reporting techniques, we analyse below the trends discernible in transparency reporting as well as the metrics that are involved.

Reporting metrics

Transparency reports demonstrate a platform's content moderation strategy and the nature of its cooperation with law enforcement and governments.

Within reporting on law and enforcement and government cooperation, platforms will typically include the law enforcement guidelines engaged, information requests, and takedown or removal requests. When content is moderated in accordance with a platform's own Terms of Service or Community Guidelines, some platforms have chosen to demonstrate user reporting numbers, proactive monitoring numbers as well as data related to user appeals. These structures are outlined below:

Government and Law Enforcement Requests:

- Information requests
- Takedown requests
- TOS report
- Removal requests under specific regulation/legislation

Content Moderation Decisions / Community Guidelines Enforcements:

- User reporting
- Proactive monitoring
- User appeals.

For a detailed view of benchmarking, please refer to the Knowledge Sharing Platform's Transparency Report Benchmarks.[278]

Trends in structure

In an analysis of the world's top 50 online content-sharing services, 11 of them are currently publishing terrorist and violent extremist content-specific transparency reports.[279] Transparency

[277] A transparency report for an encrypted messaging service cannot provide the same information as one for a content sharing platform.

[278] https://ksp.techagainstterrorism.org/knowledgebase/transparency-reporting-benchmarks/

[279] https://www.oecd-ilibrary.org/science-and-technology/transparency-reporting-on-terrorist-and-violent-extremist-content-online_8af4ab29-en;jsessionid=4CCbCWLxmhjx47olEd69mXZ0.ip-10-240-5-187

reports range greatly, with only a few platforms reporting on terrorist and violent extremist content and in doing so, each presents their metrics uniquely.

The decision of which categories and subcategories to cover is unique to each platform's resources and type.

Some platforms might report on both categories of government requests and content moderation decisions, whereas others will prefer to focus on just one category. Similarly, not all tech platforms will report on each of the sub-categories listed. Some platforms will decide to have a "step-by-step" approach to transparency reports and start by publishing a report focused on law enforcement requests or content moderation decisions only before publishing the other one. Depending on the categories and subcategories covered, certain platforms will prefer including all information in one general report, whereas others will prefer to present the information in different reports.

When publishing these reports, some companies will have the capacity to provide more details and metrics than others. The size and capacities of platforms vary greatly from one another, and not all platforms can be expected to provide the same level of detailed metrics.

All of this depends on the specificities and functionalities of each platform (for instance a platform that does not engage in proactive monitoring nor moderate comments will have no reasons to include these metrics), but also on internal capacity.

Transparency Reporting Guidelines

Tech Against Terrorism has developed the Tech Against Terrorism Guidelines on Transparency Reporting for Governments and Tech Companies.[280] These Guidelines ask tech companies to report on their policies, processes, systems, and outcomes of their counter-terrorism measures. These Guidelines incorporate essential considerations around the rule of law, proportionality, reciprocity, and smaller tech company capacity. To accompany the Guidelines, Tech Against Terrorism provides examples of transparency reporting to illustrate the recommendations.

Case studies: Platforms based in African Commonwealth Member Countries

Together with the analysis of tech sector trends in counter-terrorism policies, this report assesses the policies and practices of two tech companies in African Commonwealth member countries.

Of local platforms that originate or are headquartered in African Commonwealth member countries, most are marketplace or financial services platforms, while the social media platforms most commonly used are international in origin. Thus, the two companies this report has included for assessment are different forms of online marketplace.

For each platform, Tech Against Terrorism conducted a brief assessment of the company's policies, focusing on their prohibited conduct or content and how this is communicated to users.

All of the information gathered and analysed below has been accessed with publicly available information.

Spleet Africa

Background Information:

Type & description	Tech company (website); online marketplace for affordable residential rentals and rental financing in Africa.

[280] https://transparency.techagainstterrorism.org/

Address	https://spleet.africa/
Origin / Headquarters	Nigeria (Lagos)
Active	Nigeria, Ghana, Kenya

Policy Review:

General	▪ Easily accessible terms ▪ Contact email address provided for users ▪ Questions asked in the live chat box, acts as their support area ▪ No information found on user reporting and content moderation.[281] ▪ User reporting options not found when logged into an account or not logged in.
Terms of Service	▪ No explicit prohibition of terrorism or violent extremism However: ▪ Under “Use of the Site”: “Will not use the site for any illegal or unauthorized purposes”. ▪ Under “User generated contributions”282, amongst others, Spleet asks its users to ensure that their contributions: ○ “are not obscene, lewd, lascivious, filthy, violent, harassing, libellous, slanderous, or otherwise objectionable (as determined by us).” ○ “do not advocate the violent overthrow of any government or incite, encourage, or threaten physical harm against another.” ○ “do not violate any applicable law, regulation, or rule.” ○ “do not otherwise violate, or link to material that violates, any provision of these Terms and Conditions, or any applicable law or regulation.”

15ghana

Background Information

Type & description	Tech Company (Website), a freelance marketplace where services and products are sold by its users.
Address	https://www.15ghana.com/
Origin / Headquarters	Ghana (Accra)
Active	Ghana

Policy Review

[281] A caveat to note here is that this is not a content sharing website.

[282] According to Spleet, user generated contributions are explained as follows: “The Site may invite you to chat, contribute to, or participate in blogs, message boards, online forums, and other functionality, and may provide you with the opportunity to create, submit, post, display, transmit, perform, publish, distribute, or broadcast content and materials to us or on the Site, including but not limited to text, writings, video, audio, photographs, graphics, comments, suggestions, or personal information or other material (collectively, "Contributions").”

General	▪ Easily accessible terms ▪ Contact email address provided for users ▪ Questions asked in the live chat box, acts as their support area ▪ Support request form available,[283] which can be used to submit requests on the topics of “order support”, “review removal”, “account support”, and “report a bug”. Users can only submit a report when logged in to their account. ▪ Knowledge Bank, which acts as an FAQ’s page. For example, “why has my offer been rejected”?, which explains why some offers may not have been approved for listing.[284] ▪ Trust and Safety page[285], which offers key tools for users to use the marketplace safely, whether buying or selling services. Explains how to report problems, which applies to listings, messages and profile pages. ▪ User reporting available when logged in to the account. ▪ Customer Support: “Our Customer Support team is available 24/7 if you have any questions regarding the Site or Terms of Service.”[286]
Terms of Service	▪ No explicit prohibition of terrorism or violent extremism However: ▪ Under “Key terms”: o “ii. Jobs may be removed by 15ghana for violations to these Terms of Service, which may include (but are not limited to) the following violations and/or materials:”, includes, amongst others: ▪ “Illegal or Fraudulent services, Copyright Infringement, Trademark Infringement” ▪ “Adult oriented services, Pornographic, Inappropriate/Obscene” ▪ “Spam, Nonsense, or Violent Jobs” o “iv. Jobs that are removed for violations mentioned above, may result in the removal of the seller’s account.” o “v. Jobs that are removed for violations are not eligible to be restored or edited.” o “vi. Jobs may be removed from our Search feature due to poor performance and/or user misconduct.” o “vii. URLs in your Job text that redirect to third party websites are subject to approval and may be considered inappropriate to use on 15ghana.” o “viii. Jobs are required to have an appropriate Job image related to the service offered. An option to upload two additional Job images are available to all sellers.”

A brief analysis of the above two platforms finds that neither platform has community guidelines, all information pertaining to prohibited conduct and content is located in the Terms of Service.

There is a commendable level of customer service and general contact resources available.

There is no publicly accessible information on the platforms’ content moderation processes

[283] Support form available at: https://www.15ghana.com/customer_support
[284] Available at: https://www.15ghana.com/article/why-has-my-offer-been-rejected
[285] Available at: https://www.15ghana.com/pages/safety
[286] https://www.15ghana.com/terms_and_conditions

There is no explicit prohibition of terrorism and violent extremism, however both platforms do prohibit illegal and violent content or conduct

Dossier E: Countering terrorist use of the internet whilst respecting human rights

Executive Summary

It is well-established that counter-terrorism measures risk having an adverse impact on human rights and fundamental freedoms. Online counter-terrorism efforts are no different, and civil society groups and observers note several risks to digital rights and online freedom of expression present in both government- and company-led initiatives.

Government-led efforts can compromise human rights and freedom of expression when online activity is regulated without adequate safeguards. This includes regulation that incentivises content removal, including that of terrorist content, without appropriate checks and balances in place to examine content illegality. Observers note that this risks the removal of potentially legal or otherwise legitimate speech. Further, there are concerns that regulatory initiatives or cooperative arrangements risk undermining the rule of law, by outsourcing adjudication of content illegality to tech companies rather than to courts or judicial bodies. Lastly, civil society groups have cautioned against the use of blunt instruments such as internet shutdowns and platform blocking to tackle the spread of harmful content.

Tech company-led initiatives are criticised primarily for the potentially negative effect of automated content moderation solutions. To identify and remove terrorist content, most larger tech companies now rely on automated tools, which are effective at doing so at scale. However, such solutions often make inadequately nuanced decisions, which means that legal or otherwise legitimate content might be mistakenly identified as terrorist content. This in turn risks suppressing crucial information, including news reporting and counter-narrative campaigns. Experts also highlight the general lack of transparency from the tech industry with regards to how automated tools contribute to online counter-terrorism efforts.

Recommendations

Tech Against Terrorism recommends that governments:

- Include human rights safeguards in all counter-terrorism initiatives, including regulation and operational collaborative arrangements which target online space
- Ensure that online and offline counter-terrorism efforts do not undermine either the rule of law or principles of territoriality
- Avoid using counterterrorism as justification for indiscriminate responses, including internet shutdowns and platform blocking
- Commit to the Tech Against Terrorism Guidelines on transparency reporting on online counter-terrorism efforts for governments[287]

Tech Against Terrorism recommends that tech platforms:

- Ensure that their counter-terrorism efforts, including automated content moderation solutions, do not have impact adversely on human rights and fundamental freedoms
- Commit to the Tech Against Terrorism Pledge
- Commit to the Tech Against Terrorism Guidelines on transparency reporting on online counter-terrorism efforts for tech platforms

[287] https://transparency.techagainstterrorism.org/

Online Counterterrorism: Summary of Human Rights Risks

As governments around the world have increased their counter-terrorism operations, concerns have been raised by observers, civil society groups, and experts about the potentially negative impact on human rights such operations might have.[288]

Likewise, as collective efforts to counter terrorism online have increased in recent years, so has the critical examination of the risk such initiatives can pose to human rights and fundamental freedoms.

Below, we summarise some of the risks most frequently cited by civil society groups[289] and observers such as the UN Special Rapporteurs on human rights and counter terrorism,[290] and also report Tech Against Terrorism's own findings from practical work in this area.

Government initiatives

Below we outline the central concerns raised by experts and by Tech Against Terrorism's own research into government initiatives to counter terrorism online.

Legislation to incentivise increased content removal

As outlined in Tech Against Terrorism's Online Regulation Series,[291] the regulation of online content, including for counter-terrorism purposes, can have a negative impact on civil liberties. Such regulation includes both explicit requirements, for example to remove content, and provisions that post a more indirect risk to online speech when, out of an excess of caution, platforms are inclined towards censorship in the name of compliance.

Below are a few examples of regulatory trends capable, in the assessment of Tech Against Terrorism, of indirectly having a negative impact on human rights and fundamental freedoms:

- Tight **removal deadlines for illegal content**: such measures, especially when enforceable by financial penalties, incentivise the abandonment of due process when responding to reportedly illegal content. Such measures therefore risk the removal of arguably legal material.[292]
- **Incentives to use automated tools for content removal**: as discussed, automated solutions are useful at scale but have several drawbacks. Whether by explicitly requiring platforms to introduce automated content moderation tools or by setting targets that will be difficult for platforms to comply with them, the drive towards automation inevitably poses risks to freedom of expression
- Legal **liability for tech platforms and/or tech platform employees for third party content**: given 'platforms' aversion to the risks of litigation, it is likely that establishing platforms' liability for user-generated content on their sites will cause platforms to err on the side of removal. This naturally carries with it the risk that legal and legitimate speech will be removed.

[288] For a selection of sources, see: Human Rights, Terrorism, and Counter-terrorism: Office of the United Nations High Commissioner for Human Rights; Misuse of counter-terrorism laws threatens human rights in the OSCE region: Fair Trials, 03.10.2019; Misuse of anti-terror legislation threatens freedom of expression: Council of Europe, 04.12.2018; Human rights impact of counter-terrorism and countering (violent) extremism policies and practices on the rights of women, girls and the family - Report of the Special Rapporteur (A/HRC/46/36) [EN/AR/RU/ZH]: UN Human Rights Council, 13.02.2021; STATEMENT ON THE IMPACT OF US COUNTER TERRORISM EFFORTS IN AFRICA ON HUMAN RIGHTS BEFORE HOUSE OVERSIGHT AND REFORM NATIONAL SECURITY SUBCOMMITTEE: Amnesty International, 17.12.2019; UN: Counter-terrorism measures must uphold human rights: Article 19, 01.07.2021.

[289] https://freedomhouse.org/report/freedom-net; https://www.article19.org/resources/uk-parliament-protect-freedom-of-expression-online-and-reject-the-counter-terrorism-and-border-security-bill-2018/

[290] https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=24013&

[291] https://www.techagainstterrorism.org/wp-content/uploads/2021/07/Tech-Against-Terrorism-%E2%80%93-The-Online-Regulation-Series-%E2%80%93-The-Handbook-2021.pdf

[292] Tech Against Terrorism has warned against the potential effects of the EU regulation on terrorist content: https://www.techagainstterrorism.org/2021/06/16/tech-against-terrorism-response-to-the-eus-terrorist-content-online-regulation/

For a more detailed discussion on online regulation, see Dossier B.

Risks of undermining the rule of law

Emerging global online regulation also risks compromising the rule of law and due process.

Delegating adjudication to tech companies

A salient trend in online regulation, particularly when intended to counter terrorist content, is the outsourcing of the adjudicative function of from courts and democratically accountable institutions to private tech companies. This is inevitable when legal provisions require platforms to remove illegal content from their services (whether notified by the authorities or users or noticed in the course of proactive monitoring). In practice, such removal requirements mean that platforms themselves need to assess whether content is illegal or "harmful",[293] thus requiring tech platforms to exercise a jurisdiction properly reserved, in democratic societies, to public tribunals. This severely compromises the rule of law and due process and is contrary to international human rights standards[294] which emphasise that limits to freedom of expression should be decided by judicial authorities.[295]

Examples in Sub-Saharan Africa:

All laws that mandate platforms to remove flagged content within a short timeframe, or proactively remove certain types of content, are in effect placing the onus of adjudication of illegality on tech platforms. In Tech Against Terrorism's assessment of online regulation and legal responses to terrorist use of the internet in Commonwealth countries in Sub-Saharan Africa, this is the case with **Malawi's Electronic Transactions and Cybersecurity Act,**[296] July 2016 (came into force in 2017) which requires platforms to inform the country's Communication Regulatory Authority of illegal content reported on its services.

For more examples of regulations leading to the outsourcing of the adjudication of legality to tech platforms globally, please see The Online Regulation Series Handbook (Section 1 on Key Trends in online regulation).

Broad definitions

Another trend in emerging online regulation which presents risks for the rule of law, is the use of vague or broad definitions of targeted content categories. This is particularly true of legislation targeting "harmful online content" which is often impractically broad in its definition of what constitutes harmful content, with minimal indication of how to operationalise these definitions to assess and action content correctly. Definitions of terrorist content are no exception to this and sometimes border on the circular – as is the case in the draft UK Online Safety Bill, which defines terrorist content as content that leads to a terrorist offence.[297]

Vague definitions of such foundational concepts make it significantly difficult for platforms to comply with legislation by implementing appropriate and proportionate moderation measures. Vague definitions also entail significant risks to freedom of expression when platforms err on the

[293] Some jurisdictions, such as the United Kingdom, are considering the introduction of legislation which would compel companies to remove legal but "harmful" content. See our Online Regulation Handbook: https://www.techagainstterrorism.org/wp-content/uploads/2021/07/Tech-Against-Terrorism-%E2%80%93-The-Online-Regulation-Series-%E2%80%93-The-Handbook-2021.pdf

[294] https://www.undocs.org/A/HRC/38/35

[295] This is exemplified by the criticism made by David Kaye of the French "cyberhate law". The law itself did not create a new set of harms (it was based on restrictions to freedom of expression existing in French law); nevertheless, Kaye underlined that "censorship measures", such as those implied by the duty to remove terrorist and hateful content, should not be delegated to private entities.

[296] file:///Users/maygane.janin/Downloads/num_act_2016_33%20(1).pdf

[297] See the section on the United Kingdom in the Online Regulation Series Handbook: https://www.techagainstterrorism.org/2021/07/16/the-online-regulation-series-the-handbook/

side of caution to avoid penalties and opt for broader interpretations of legal provisions to be on the safe side, and thus expose legal or otherwise legitimate speech to the risk of removal.

With such vague definitions of "legal but harmful" content, countries risk vitiating freedom of expression when regulations place commercial pressure on tech companies to remove legal or non-violent speech. This is particularly the case when online regulation extends beyond illegal content to cover for "harmful" content, which raises the prospect of criminalising online speech that is legal offline. Whilst tech companies are free to act in a way that protects the interests of users, including by removing hate speech (which might be legal), it sets a potentially dangerous precedent if governments either directly or indirectly pressure tech companies to remove legal speech.

The risks of "censorship creep"

Danielle Citron (Professor at the University of Virginia School of Law and expert on information privacy and free expression), in her criticisms of the EU regulation of online content and EU Internet Referral Units, has expressed concerns with the risks of "censorship creep[298] whereby a wide array of protected speech, including political criticism and newsworthy content, may end up being removed from online platforms on a global scale." Citron's criticisms focus on "definitional ambiguity" around what constitutes harmful content, namely "hateful conduct" and "violent extremism material", which can be abused to target legitimate speech and political dissent, as well as to pressure platforms to over-remove content.

Extraterritorial implications

Certain regulations intended to counter harmful content also raise the broader question of extraterritorial jurisdiction, as several laws and legislative initiatives introduced in the last four years violate the principles of territoriality by requiring removal of content across platforms beyond their native jurisdictions. The legislation to regulate the online space currently under consideration in Brazil, for example, specifies no territorial limit to its application and could therefore apply globally.[299] The 2020 Citizens Protection Rules, introduced in Pakistan in October 2020, were also criticised for applying to all content published by Pakistani citizens outside of the country's jurisdiction.[300]

Comparable to such legislative mandates in the concern they elicit are judicial rulings which enjoin tech companies to apply a removal or banning order worldwide rather than to simply block access for local users. In August 2020, Facebook complied with an injunction issued by a Brazilian judge to block the accounts of 12 of President Bolsonaro's supporters worldwide.[301] Facebook stated that it complied with the order due to the threat of criminal liability faced by one of its employees.[302] In 2020, Facebook also had to comply with a court order to remove worldwide references to defamatory comments made against an Austrian politician.[303]

Governments may have speech codes and limits to freedom of expression nationally, but they should not be enforced beyond their jurisdictional limits, when to do so would be to interfere with standards of acceptable speech in other countries.

Extra-legal pressure on platforms

Related to the outsourcing of the adjudication of illegality to tech platforms is the question of how governments use extra-legal bodies and norms to pressure tech companies to remove content. This

[298] https://scholarship.law.nd.edu/ndlr/vol93/iss3/3/

[299] https://www.techagainstterrorism.org/2020/11/11/the-online-regulation-series-brazil/

[300] https://www.techagainstterrorism.org/2020/10/06/online-regulation-month-pakistan/; https://www.eff.org/deeplinks/2016/08/global-ambitions-pakistans-new-cyber-crime-act

[301] The individuals were under investigation for running a fake news network.

[302] This is not the first time that a Facebook employee has faced criminal liability in Brazil for corporate non-compliance with a court order. In 2016, Diego Dzodan, Facebook's Vice President for Latin America was jailed for 24 hours, following a disputed court order for WhatsApp to disclose user data for a drug-trafficking investigation.

[303] Compliance will be required as long as the injunction remains in force. https://techcrunch.com/2020/11/12/facebook-loses-final-appeal-in-defamation-takedown-case-must-remove-same-and-similar-hate-posts-globally/

problem is particularly acute when clear regulation mandating platforms to remove certain types of content is unavailable.

Exemplifying this are Internet Referral Units (IRUs). These are law enforcement bodies operating within national or regional police structures and which report suspected illegal content, notably terrorist content, to tech companies for assessment and takedown as a breach of the company's Terms of Service. Contrary to removal requests, which can be sent by judicial authorities to compel platforms to remove illegal content, IRUs enforce platforms' own guidelines about what is acceptable content on their services, rather than existing national laws. The UK, France, the Netherlands, and Europol all have such units.[304]

For more a more in-depth analysis of IRUs, please see the dedicated section (4.b) in Dossier C Public-private partnerships and cross-industry cooperation.

Beside IRUs, there is the question of governments pressuring tech companies to remove legal content by exploiting broad definitions of harmful content (see above), including of terrorist content and of speech deemed to present risks to national security.

Automated solutions and the rule of law

Very little online regulation passed in recent years explicitly mandates platforms to monitor their services proactively to detect and remove illegal or harmful content.[305] However, stringent requirements to act against and swiftly remove terrorist content from online platforms will, in practice, require a significant increase in resources dedicated to content moderation as well as an increased reliance on automated content moderation tools, in particular to detect and prevent the upload (or livestream) of terrorist and other illegal content/

The draft UK Online Safety Bill is an example of regulation which incentivises the use of automated moderation solutions without explicitly requiring or referencing it. Amongst the different duties of care imposed on tech platforms by the draft bill, the provisions relating to mitigating and managing risk require platforms to deploy "proportionate systems and processes" to counter illegal content – including to "swiftly take down such content".[306]

Whilst automated content moderation has its benefits, current solutions are insufficiently nuanced to comprehend user-generated content and correctly assess whether certain pieces of content are in fact terrorist material or harmful. See more on this in "B) Tech company initiatives" below.

The deployment of automated solutions to detect and remove terrorist content faces an immediate practical obstacle. These solutions cannot formulate or supplant consensus on what constitutes a terrorist organisation and need to be informed by relevant national and international proscriptions. Such deployment is especially complicated when harmful content originates from users that are not officially affiliated with a terrorist or violent extremist outfit, or when the content is confined to a legal "grey area" without further examination

Operational measures

Internet shutdowns

Internet shutdowns are a tactic used increasingly by governments to restrict speech and disrupt online communications during times of unrest. Early noteworthy cases of countrywide internet

[304] See Dossier C: Public-private partnerships and cross industry cooperation, and also https://globalnetworkinitiative.org/human-rights-risks-irus-eu/

[305] Amongst the 17 jurisdictions analysed by Tech Against Terrorism in the Online Regulation Series Handbook, only India and Pakistan included explicit provisions to conduct pro-active monitoring or use automated moderation solutions, the EU having scrapped the provisions related to upload filters from its final version of the regulation on preventing the dissemination of terrorist content online.
See: https://www.techagainstterrorism.org/2021/07/16/the-online-regulation-series-the-handbook/; https://edri.org/our-work/terrorist-content-regulation-document-pool/

[306] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safety_Bill_Bookmarked.pdf

shutdowns occurred in the Maldives (2004),[307] Guinea (2007)[308], Iran (2009)[309], Tunisia (2010), and Egypt (2011)[310], whereas local mobile service was shut down on parts of the San Francisco subway line in 2011.[311]

Internet shutdowns constitute a disproportionate suppression of speech, and on that ground have been criticised by numerous international and regional bodies, including the United Nations, African Commission on Human and Peoples' Rights, the Organization of American States, the Organization for Security and Co-operation in Europe, the Freedom Online Coalition, and multiple civil society organisations.[312].

Despite this, the method has been employed with increasing frequency. While data prior to 2016 is patchy, and the true total number of internet shutdowns is likely unknowable, nearly 850 intentional shutdowns have been documented and verified over the last decade via Access Now's Shutdown Tracker Optimization Project (STOP). Of these, 768 shutdowns across 63 countries have taken place just in the last five years. There were 196 global shutdowns implemented by 25 countries in 2018, 213 (in 33 countries) in 2019, and 155 (in 29 countries) in 2020. In 2020, there were 18 shutdowns in Africa carried out by 10 different countries.[313] Between January and May of 2021, there were 50 shutdowns implemented by 21 different countries.

Whilst justifications from governments vary – Access Now has proven that governments seek to control the information ecosystem in times of political instability, protests, elections, or significant holidays – terrorism is occasionally cited as a justification for closing down the internet. According to Access Now's collection of internet shutdown data, national security and counter-terrorism purposes have been cited as justifications in 10 shutdowns in Myanmar and India.[314]

In addition to shutdowns being a disproportionate response to terrorism – research shows that shutdowns can severely disrupt vital infrastructure and health care systems – counter-terrorism experts and digital rights advocates have highlighted that it is unlikely to be effective in preventing terrorism and terrorist use of the internet. As noted in a Global Network on Extremism & Technology study:

"[B]lanket internet shutdowns fail to acknowledge the power of word of mouth and the ability of people eager to mobilise and express dissent to use VPN servers or peer-to peer chat services relying on Bluetooth connectivity to circumvent these shutdowns […] extremists intended to spread hate can also access VPNs and circumvent these bans"[315]

There is also some research that suggests that such measures may lead to an increase in violent as opposed to non-violent protests.[316]

Platform and/or website blocking

Platform blocking has also been implemented by governments as a way to control or restrict online communications. Noteworthy cases include Russia's blocking of encrypted messaging app Telegram in 2018 (which has since been lifted), after the platform refused to give the Russian government

[307] http://www.ipsnews.net/2004/08/rights-maldives-unrest-worries-international-community/

[308] Admire Mare, "State-Ordered Internet Shutdowns and Digital Authoritarianism in Zimbabwe," International Journal of Communication, no. 14 (2020): 4244–63

[309] https://www.cbsnews.com/news/iran-blocks-internet-on-eve-of-rallies/

[310] Jim Cowie, "Egypt Leaves the Internet," Internet Intelligence (blog), January 27, 2011, https://blogs.oracle.com/internetintelligence/egypt-leaves-the-internet-v3

[311] https://jigsaw.google.com/the-current/shutdown/#references

[312] https://www.oas.org/en/iachr/expression/showarticle.asp?artID=1146&lID=1

[313] https://www.accessnow.org/cms/assets/uploads/2021/03/KeepItOn-report-on-the-2020-data_Mar-2021_3.pdf

[314] https://www.orfonline.org/expert-speak/kashmir-blackout-counter-terrorism-increasingly-challenging-role-internet-55119/

[315] https://gnet-research.org/2020/07/31/from-fears-to-conviction-why-internet-shutdowns-dont-work/

[316] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3330413

access to its decryption keys, in violation of the country's anti-terrorism laws.[317] Iran blocked Telegram in 2018, citing 'national security'.[318] This followed Indonesia's 2017 blocking of Telegram, which was justified by the fact that terrorist groups were using the app.[319] Egypt blocked 21 websites in 2017 for counter-terrorism purposes, including news websites like al-Jazeera and Mada Masr.[320] More recently, India banned social media platform TikTok for alleged disruption of public order.[321]

Other countries are considering introducing legal mechanisms to block platforms. In Australia, the new Online Safety Act[322] allows the government to block apps if they are found to "facilitate" the posting of specific illegal material, including terrorist material.[323]

Tech company initiatives

Observers and activists have also noted risks to human rights in tech platforms' response to terrorist use of their platforms, most of which concern the potentially negative impact of automated content moderation solutions. We provide examples of the central concerns below.

Automated removal tools

Content moderation decisions can have significant ramifications for human rights, including freedom of expression but also associated freedoms, particularly given the role that online service providers play in public discourse. The risk for vulnerable or marginalized communities is especially significant given the amenability of counter-terrorism laws to the suppression of online dissent amongst civilian populations.

Much online content, and extremist content in particular, is moderated increasingly by automated flagging and tools. In this process, platforms use automated tools to detect potentially violative content which is then submitted to a human moderator for review - often before content is seen by users.[324] In the dossier entitled "Tech Sector Efforts to Counter-terrorism", we discuss some commonly used automated tools for content moderation, including digital hash technology and image recognition. In doing so, we elucidate the important limitations of such automated tools, including: the negative impact on freedom of speech when legitimate speech or content is removed under counter-terrorism policies; unwarranted surveillance; lack of transparency and accountability in the development process which prevents scrutiny by external reviewers; and the accidental deletion of digital evidence.[325]

Below, Tech Against Terrorism outlines the risks to freedom of expression inherent in automated tools given their failure to account for context, the possibility of false positives, and the likelihood of unintended and adverse consequences, such as the removal of legitimate and legal content and digital evidence. For more information on this, please see Tech Against Terrorism's report entitled "Gap Analysis and Recommendations for deploying technical solutions to tackle the terrorist use of the internet".[326]

[317] https://www.theverge.com/2020/6/18/21295535/russia-telegram-ban-lifted-security
[318] https://www.bbc.com/news/technology-43963927
[319] https://www.reuters.com/article/us-indonesia-security-apps-idUSKBN1AH40K
[320] https://www.reuters.com/article/us-egypt-censorship-idUSKBN18K307
[321] https://pib.gov.in/PressReleseDetailm.aspx?PRID=1635206#
[322] https://www.communications.gov.au/have-your-say/consultation-bill-new-online-safety-act
[323] For more information, please see our Australia entry in the Online Regulation Series Handbook: https://www.techagainstterrorism.org/2021/07/16/the-online-regulation-series-the-handbook/
[324] https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#Introduction
[325]https://gifct.org/wp-content/uploads/2021/07/GIFCT-TAWG-2021.pdf?utm_source=Tech+Against+Terrorism&utm_campaign=dad15fffda-EMAIL_CAMPAIGN_2019_03_24_07_51_COPY_01&utm_medium=email&utm_term=0_cb464fdb7d-dad15fffda-141408947
[326] https://gifct.org/wp-content/uploads/2021/07/GIFCT-TAWG-2021.pdf

Failure to account for context

Automated tools are unable to comprehend the nuances and contextual variations of human speech.[2] This includes their only limited ability to parse and understand variances in language and behaviour based in different demographic and regional factors. Automated tools are also limited in their ability to learn from content and can quickly become outdated.[327]

Criticism of Algorithmic Tools
The New America Open Technology Institute published 'Everything in Moderation
An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User-Generated Content'. The report found that on Twitter, members of the LGBTQ+ community found that there was a significant lack of search results that incorporated hashtags such as #gay and #bisexual, which created suspicions of censorship. The company stated that this absence was due to the deployment of an outdated algorithm that mistakenly identified posts with these hashtags as potentially offensive.[328]

The example above emphasises the need to continuously update algorithmic tools. When human reviewers engage in content moderation, they are able to make educated inferences about the meaning of speech by drawing on additional information about the case, such as the identity of the user accused of violating the platform's rules. However, to make such inferences and procedures available to an automated tool risks entrenching biases around particular constituencies of users that result in the skewed or even discriminatory enforcement of content policies.[329]

Extremist content and hate speech is characterised by a range of nuanced variations in speech related to different groups and regions, and this context can be critical in understanding whether or not it should be removed. For example, to circumvent moderation, some hateful groups have used different representations for indicating hate. A past case of this was seen when white supremacists used the names of companies, such as "Google" and "Yahoo" to replace ethnic slurs.[330] It is therefore difficult to develop comprehensive datasets for these categories of content, and equally difficult to develop and operationalize a tool capable of reliable application to different groups and regions of users, and to their variable habits of speech.[331]

False positives and unintended consequences

This inability of automated tools to account for context, as well as their tendency towards inaccuracy and unreliability, can result in false positives as well as other unintended consequences, including the reflexive censorship of specific forms of content and of specific groups. Below Tech Against Terrorism outlines the consequences of this.

Content: Social media documentation of human rights violations and crimes against humanity is critical to the effort to deliver justice and accountability. Videos and text posted online might provide the only evidence that such a monumental breach has been committed.[332] However,

327https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/the-limitations-of-automated-tools-in-content-moderation
328shttps://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/the-limitations-of-automated-tools-in-content-moderation
Hillary K. Grigonis, "Social (Net)Work: What can A.I. Catch — and Where Does It Fail Miserably?," Digital Trends, February 3, 2018, https://www.digitaltrends.com/social-media/social-media-moderation-and-ai/.
329https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/the-limitations-of-automated-tools-in-content-moderation/
330 https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/the-limitations-of-automated-tools-in-content-moderation/
331https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/the-limitations-of-automated-tools-in-content-moderation/
332 https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#Introduction

automated tooling in content moderation can often result in its deletion. This is doubly concerning given that, according to the Electronic Frontier Foundation, “restoring wrongfully deleted content is nearly impossible if the person who posted the content is not alive, is arrested, or does not have access to email, all common issues in conflict zones”.[333] Below Tech Against Terrorism outlines its analysis of the effect of YouTube’s content moderation on the documentation of conflict in Syria, Yemen, and the Ukraine.

Example: Syria

In Syria, human rights defenders have used social media platforms to log violations in the course of the conflict, and to both publish and publicise their records, and have done so effectively and often. In 2019, according to the Electronic Frontier Foundation, there were “more hours of social media content about the Syrian conflict than there have been hours in the conflict itself”.[334]

The Electronic Frontier Foundation noted that YouTube had used automated flagging tools powered by machine learning to terminate thousands of Syrian YouTube channels that were publishing videos of human rights violations. Amongst the YouTube channels that were removed, the Syrian Observatory for Human Rights, the Violation Documentation Center, Sham News Agency, and the Aleppo Media Center.[335] These cases demonstrate that the accounts suppressed included those presenting footage of protests in Syria as well as non-traditional media reporting on violent acts, none of which purported or could reasonably be said to incite violence or to encourage harmful behaviour.[336]

“At least 206,077 videos documenting rights violations were made unavailable on YouTube between 2011 and May 2019. This includes 381 videos documenting airstrikes that targeted hospitals or medical facilities.”[337]

Example: Yemen & Ukraine

The Electronic Frontier Foundation has noted similar examples in Yemen and the Ukraine. In Yemen, the war since 2015 has led to tens of thousands being killed and millions displaced. YouTube videos depicting some of the crimes and atrocities have been made unavailable. In addition, footage documenting elements of the conflict of great geopolitical significance and critical to improving worldwide understanding, such as footage showing the arming of pro-Russia and anti-government forces, has been removed.[338]

Groups: Discriminatory treatment of vulnerable and minority groups’ online speech

From the above analysis, it is clear that it is very difficult for automated tools, and even for human reviewers, to consistently differentiate activism, counter-speech, and satire about extremism from extremism itself.[339] The Electronic Frontier Foundation states that ”blunt content moderation

[333] https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#Introduction

[334] https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#Introduction

[335]https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#Introduction

[336] https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#Introduction

[337] https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#Introduction

[338] https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#Introduction

[339] https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#Introduction

systems at scale inevitably make mistakes, and marginalized users are the ones who pay for those mistakes".[340]

In addition to issues of context, the presence of bias in automated tools risks further marginalising and censoring groups that already face disproportionate prejudice and discrimination online and offline.[341] An example of such discrimination can be seen in Natural Language Processing (NLP) tools, which are a set of techniques that use software to parse text. In the context of content detection and moderation, text is parsed, or analysed with respect to its grammatical components, in order to make predictions about the meaning of the text, such as what sentiment it might indicate.[342] These tools are typically used to parse text in English. Therefore, such tools have a lower accuracy when parsing non-English text or are unable to process certain differences in dialect and language use, because they have a less substantial basis in prior learning and data to inform their analysis. This can result in harmful outcomes for non-English speakers, especially when it is applied to languages that are less prominent online – which is a deficiency particularly devastating to online counter-terrorism efforts.[343]

An example of such a community being impacted by content moderation can be seen in an instance from 2017, when a Facebook group advocating for the independence of the Chechen Republic of Iskeria, called for "Independence for Chechnya!", and was mistakenly removed for violating the company's community standards which prohibit "organizations engaged in terrorist activity or organized criminal activity." The removal was carried out despite the fact that training manuals for content moderators specifically identify causes positively associated with the Chechen Republic of Iskeria as "not violating" the rules. A Facebook spokesperson said that the deletion was made in error.[344]

'Content cartels'

Online speech expert Evelyn Douek has cautioned against what she calls "content cartels". Douek uses this term to describe tech industry collaborative arrangements which work to remove illegal and harmful content online, including child sexual abuse material and terrorist content.[345] One example mentioned by Douek is the hash-sharing database maintained by the Global Internet Forum to Counter Terrorism (discussed at length in Dossier D), in which participating companies share hashes of terrorist content. Hashes are digital 'fingerprints' of specific imagery, and the hash-sharing database allows for the fast identification of content which matches material that has previously been hashed. This means that platforms using the database can automatically identify a terrorist video so long as this content has been identified and hashed previously. Currently 17 tech companies – including Facebook, YouTube, and Twitter – participate in this scheme.[346] The majority of content flagged is removed automatically. Since participating platforms can all submit hashes to the database, this in theory means that content hashed by Twitter will automatically be removed from Facebook. Douek warns that arrangements like these lead to unaccountable and non-transparent content moderation, and that it is unclear what rules define which content gets hashed and what safeguards are in place to ensure legal or otherwise legitimate content is not removed.[347]

[340] https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#Introduction
[341] Duarte, Llansó, and Loup, Mixed Messages?
[342] https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/how-automated-tools-are-used-in-the-content-moderation-process
[343] https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/the-limitations-of-automated-tools-in-content-moderation
[344] https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#Example1
[345] https://knightcolumbia.org/content/the-rise-of-content-cartels
[346] https://gifct.org/wp-content/uploads/2021/07/GIFCT-TransparencyReport2021.pdf
[347] https://knightcolumbia.org/content/the-rise-of-content-cartels

Similar concerns about the GIFCT hash-sharing database have also been raised by public interest groups.[348]

Possible mitigation strategies

Alternatives to content removal:

In the dossier entitled “Tech Sector Efforts to Counter-terrorism”, Tech Against Terrorism provides a list of strategies commonly employed by tech companies to moderate content short of the wholesale removal of content or accounts. The aim of this list is to inform platforms, in particular small ones, seeking to understand or establish procedures which satisfy the requirements to demote or delete harmful content and at the same time to preserve freedom of speech. The alternatives to removal include the following:

- **Hiding content** either partially or for targeted sets of users is a way of removing content that some users find offensive which avoids suppressing potentially legal, legitimate, and/or permissible content across the entire platform. For example, this might be for users from a more “vulnerable demographic” or from a country where the content violates laws (but does not in others).
- **Disengagement schemes** are a means of suppressing content/users by decreasing engagement or activity around a post or account without actually removing them from the platform entirely. This can be done by manipulating a site’s functionality to not promote such content further, or by demoting users that might otherwise be given a favoured platform. Disengagement is distinct from hiding content since disengagement demote a piece of content or users’ prominence across the whole platform, rather than only for certain sections of users.
- **Educational or comms-based tactics** seek to provide users with extra information about a piece of content so that they can decide for themselves whether to see it or engage with it, thereby relieving some of the onus on the tech company. Such strategies can be understood as a means of empowering users, but they do still rely on intervention by a platform to adjudicate whether further information should be provided and what that information should be.
- **Features focused on community empowerment** rely on users themselves to curate the online space that they want to see. These strategies are ubiquitous among community-reliant platforms, and among platforms that have smaller teams of content moderators. Such strategies are particularly relevant for content that is offensive but does not strictly violate a platform’s content standards.

It is important to note that none of the content moderation measures mentioned above should be seen as a panacea in the struggle against violent extremist or terrorist content online. Companies will most likely have to employ more than one method of content moderation to maintain the efficacy of their overall strategy, and different methods will inevitably work better on different types of platforms. Ultimately, any content moderation strategy must be based on clearly enforceable content standards to preserve natural justice, and the rights of users.

Finally, in line with the principles outlined in the Tech Against Terrorism Pledge,[349] any moderation or removal process should be accompanied by accessible and expedient mechanisms for appeal and redress to ensure that users are able to contest any decision made by the platform. Moreover, Tech Against Terrorism urges tech companies, especially those of larger size, to include in their transparency reports available data on user and content removal for violating terms of service and community guidelines.

[348] https://www.voxpol.eu/one-database-to-rule-them-all/
[349] https://www.techagainstterrorism.org/membership/pledge/

Internet shutdowns

According to open-source reporting, there have been government-ordered internet shutdowns in areas within the following sub-Saharan Commonwealth member countries since 2017:

Country	Reason(s) provided
Tanzania[350]	Counterterrorism
Uganda[351]	Election integrity
Cameroon[352]	Political unrest

In 2020, Kenya also experienced two network disruptions in Mandera County as a result of Al Shabaab attacks.

Platform blocking

According to open-source reporting, specific platforms have been blocked in the following African Commonwealth member countries:

Country	Platform(s)
Uganda[353]	Facebook
Zambia[354]	Facebook, Messenger, Instagram, Twitter, WhatsApp
Nigeria	Twitter

Case study: Nigeria and Twitter

On 4 June 2021, Nigeria's Ministry of Information and Culture announced the indefinite suspension of Twitter services in the country, stating the presence of activities on the social media platform "capable of undermining Nigeria's corporate existence".[355]

This shutdown of Twitter services has been criticised for violating Nigerians' rights to freedom of expression and access to information,[356] and is estimated to affect 104.4 million users in Nigeria and to have cost the country around $366.9 million (as of August 2021, with an estimated loss of $6 million per day).[357]

As of 5 September 2021, the suspension remains in place despite an earlier announcement in August 2021 by Information Minister, Lai Mohammed, that an "amicable resolution is very much in sight" and that the suspension would soon be lifted.[358]

[350] https://www.dw.com/en/tanzania-internet-slowdown-comes-at-a-high-cost/a-55512732
[351] https://www.accessnow.org/who-is-shutting-down-the-internet-in-2021/#facts-figures
[352] https://techtribes.org/cameroon-shuts-down-the-internet-for-240-days/
[353] https://www.nytimes.com/2021/01/13/world/africa/uganda-facebook-ban-elections.html
[354] https://netblocks.org/reports/whatsapp-and-social-media-restricted-in-zambia-on-election-day-18lpLY8a
[355] https://twitter.com/FMICNigeria/status/1400843067062734858
[356] https://www.accessnow.org/nigeria-blocks-twitter-keepiton/; https://www.accessnow.org/nigeria-twitter-ban-ecowas-court/; https://www.article19.org/resources/nigeria-authorities-must-stop-clamping-down-on-digital-rights/
[357] https://qz.com/africa/2043666/twitter-ban-has-cost-nigeria-over-360-million-in-two-months/
[358] https://www.voanews.com/africa/nigerian-minister-says-twitter-ban-be-lifted-soon

The Community Court of Justice of the Economic Community of West African States (ECOWAS) is to hear arguments challenging the ban on 29 September 2021.[359]

A complete timeline of Nigeria's decision to suspend Twitter can be found at the end of this case study.

Analysts assert that Nigeria's Twitter ban results from Twitter's decision on 2 June 2021 to remove two tweets from President Muhammadu Buhari threatening Southern separatists by referring to the country's past civil war,[360] which Twitter assessed to be in violation of its policy on abusive behaviour.[361] Alongside the removal, the President's account was also suspended by Twitter for 12 hours.[362] Despite the close succession of events, the Nigerian government has denied that the decision to ban Twitter was motivated by the platform's enforcement of its moderation policies. Instead, the government has been arguing that the ban was linked to misinformation and "activities" on Twitter which had "real world violent consequences"[363] and were threatening the unity of the country.[364]

Facebook: Facebook took comparable action against the same post by President Buhari about the Southern separatists, but the Nigerian government has taken no action against the social media platform. According to 'Rest of World', contrary to Twitter, Facebook had reached out to the Nigerian government to edit the post before it was removed – a request rejected by the Nigerian government.[365]

Following the suspension, the Nigerian government instructed the National Broadcast Commission (NBC) to begin the process for licensing online platforms ("Over-The-Top (OTT) media services" and social media platforms).[366] On 10 June 2021, the NBC asked all online platforms to apply for a broadcast licence if they wished to continue operating in the country. Subsequently, the Minister of Information and Culture, Lai Mohammed, asked Nigerian lawmakers to amend the National Broadcast Act so as to empower the NBC with the regulation of online platforms.

Background to the Twitter Ban

Nigeria's decision to suspend Twitter needs to be understood in the context of increased efforts by the Government to regulate social media, and of the importance of Twitter in social protests in the country.

Commentators that are analysing the 2021 suspension of Twitter note that it occurred within a few months of the #EndSARS protest movement of October 2020.[367]

At the time of the October 2020 protests, the Nigerian government had already attempted to strengthen regulation of social media use by reviving and strengthening its campaign to pass its so-called "social media" bill, the "Protection from internet falsehood and manipulation bill and other related matters" (introduced in 2019)[368] which aims to tackle the spread of misinformation that has

[359] https://www.mediadefence.org/news/ecowas-court-adjourns-nigeria-twitter-ban-case/

[360] https://www.aljazeera.com/news/2021/6/8/nigerians-launch-legal-action-against-governments-twitter-ban
https://www.france24.com/en/live-news/20210607-nigeria-says-in-talks-with-twitter-after-suspension

[361] https://www.article19.org/resources/nigeria-authorities-must-stop-clamping-down-on-digital-rights/

[362] https://freedomhouse.org/article/nigerias-twitter-ban-bellwether-case-internet-freedom

[363] https://www.france24.com/en/live-news/20210607-nigeria-says-in-talks-with-twitter-after-suspension

[364] https://twitter.com/FMICNigeria/status/1400843067062734858

[365] https://restofworld.org/2021/inside-nigerias-decision-to-ban-twitter/

[366] https://guardian.ng/saturday-magazine/cover/social-media-regulation-between-failed-attempts-and-buharis-current-move/

[367] Before the #EndSARS protests, Twitter was already a major platform for exchanging and discussing information in the country, and to conduct social media campaigns, as exemplified by the #BringBackOurGirls campaign in 2014 after Boko Haram kidnapped 76 schoolgirls.
See: https://www.france24.com/en/live-news/20210607-nigeria-says-in-talks-with-twitter-after-suspension; https://freedomhouse.org/article/nigerias-twitter-ban-bellwether-case-internet-freedom

[368] The bill first emerged in 2015 as "A Bill for an Act to Prohibit Frivolous Petitions and other Matters Connected therewith", including a provision on death penalty for "hate speech." At the time, the bill passed

"threatened the unity of the country".[369] If passed, the bill would prohibit users from sharing certain types of content online, including content that can "Affect the security of any parts of Nigeria." Civil society organisations have raised concerns that the bill can be used to silence political dissent and non-violent political speech.[370]

In attempting to regulate online space and social media platforms, Nigeria follows a trend identified by Tech Against Terrorism in the Online Regulation series Handbook. In arguing that a social media bill – and in the case of Twitter, the shutdown of the services – is needed to counter misinformation and its "real world" consequences, Nigeria joins Singapore, India, and Brazil in a group of countries which all have passed or introduced regulatory proposals to counter the spread of misinformation, often via correction and removal orders to be issued by the government.[371]

Content moderation and the understanding of local context:

Beside the question of countering misinformation, the issue of social media platforms' understanding of the local context also appears to be at the core of the suspension of Twitter and attempts to regulate social media in Nigeria. Rest of World's reporting on the decision to ban Twitter stressed that the decision was underpinned by a "growing consensus within the government calling on Twitter to establish a local presence in order to grasp local context."[372] In relation to this, media reports have stressed that the Nigerian government saw "as a snub" Twitter's announcement in April 2021 to open its first African office in Ghana rather than in Lagos, Nigeria's economic capital.[373] In line with his argument for greater understanding of local context, and as a consequence of the decision to ban, it appears that Information Minister, Lai Mohammed, is requesting that social media platforms hire local teams and secure a licence to operate in the country.[374]

However, Freedom House alleges that despite the government's criticism of Twitter's lack of understanding of Nigeria's context, the decision to remove a tweet in reference to the country's 1967-70 civil war does in fact show some understanding and sensitivity of the country's history.

Effects on human rights

The suspension of Twitter has been criticised for its negative impact on the fundamental rights of Nigeria citizens, and specifically their rights to freedom of expression and access to information. These criticisms were heightened due to the significant use of Twitter by Nigerians to criticise the government and share critical evidence of police brutality during the #EndSARS protests in 2020.[375]

Civil society organisations have also criticised the suspension for violating the country's 1999 constitution, which recognises the rights to freedom of expression and to access to information,[376] as well as several international standards.[377] The suspension is also in violation of different regional standards on fundamental rights, including the African Charter on Human and People's Rights and

the second reading in Parliament, but President Buhari distanced himself from it due to human rights concerns. Lawmakers eventually withdrew the bill. See: https://www.theafricareport.com/51915/nigeria-social-media-bill-threatens-death-penalty-for-hate-speech/
The 2019 version of the bill is said to have plagiarised the Protection from Online Falsehood and Manipulation Act (POFMA) passed in Singapore in 2019. See: https://techpoint.africa/2019/11/28/nigerias-social-media-bill/

[369] https://www.theafricareport.com/51915/nigeria-social-media-bill-threatens-death-penalty-for-hate-speech/
https://qz.com/afrideca/1926334/endsars-nigerian-government-looks-to-regulate-social-media/

[370] https://www.amnesty.org/en/latest/news/2019/12/nigeria-bills-on-hate-speech-and-social-media-are-dangerous-attacks-on-freedom-of-expression/

[371] https://www.techagainstterrorism.org/2021/07/16/the-online-regulation-series-the-handbook/

[372] https://restofworld.org/2021/inside-nigerias-decision-to-ban-twitter/

[373] It should be noted that the opening of an office in Ghana was accompanied by Twitter advertising for jobs focused on covering Nigeria. https://restofworld.org/2021/inside-nigerias-decision-to-ban-twitter/

[374] https://restofworld.org/2021/inside-nigerias-decision-to-ban-twitter/

[375] https://freedomhouse.org/article/nigerias-twitter-ban-bellwether-case-internet-freedom

[376] https://www.article19.org/resources/nigeria-authorities-must-stop-clamping-down-on-digital-rights/

[377] Including the International Covenant on Civil and Political Rights, the International Covenant on Economic Social and Cultural Rights.
See: https://www.article19.org/resources/nigeria-authorities-must-stop-clamping-down-on-digital-rights/

the 2019 Declaration of Principles on Freedom of Expression and Access to Information in Africa which stipulates that “States shall facilitate the rights to freedom of expression and access to information online and the means necessary to exercise these rights.” (37th principle).

Internet shutdown measures, whether general or provider-specific, as in the case of Twitter, have often been criticised for being disproportionate to their stated aims in addition to their negative impact on fundamental rights. In June 2011, the UN Special Rapporteur on Freedom of Expression, the Organisation for Security and Cooperation in Europe, the Organisation of American States and the African Commission on Human and People’s Rights reaffirmed that shutdown measures are not legitimate unless they are provided for in law, in pursuance of a legitimate aim, and necessary in a democratic society.[378]

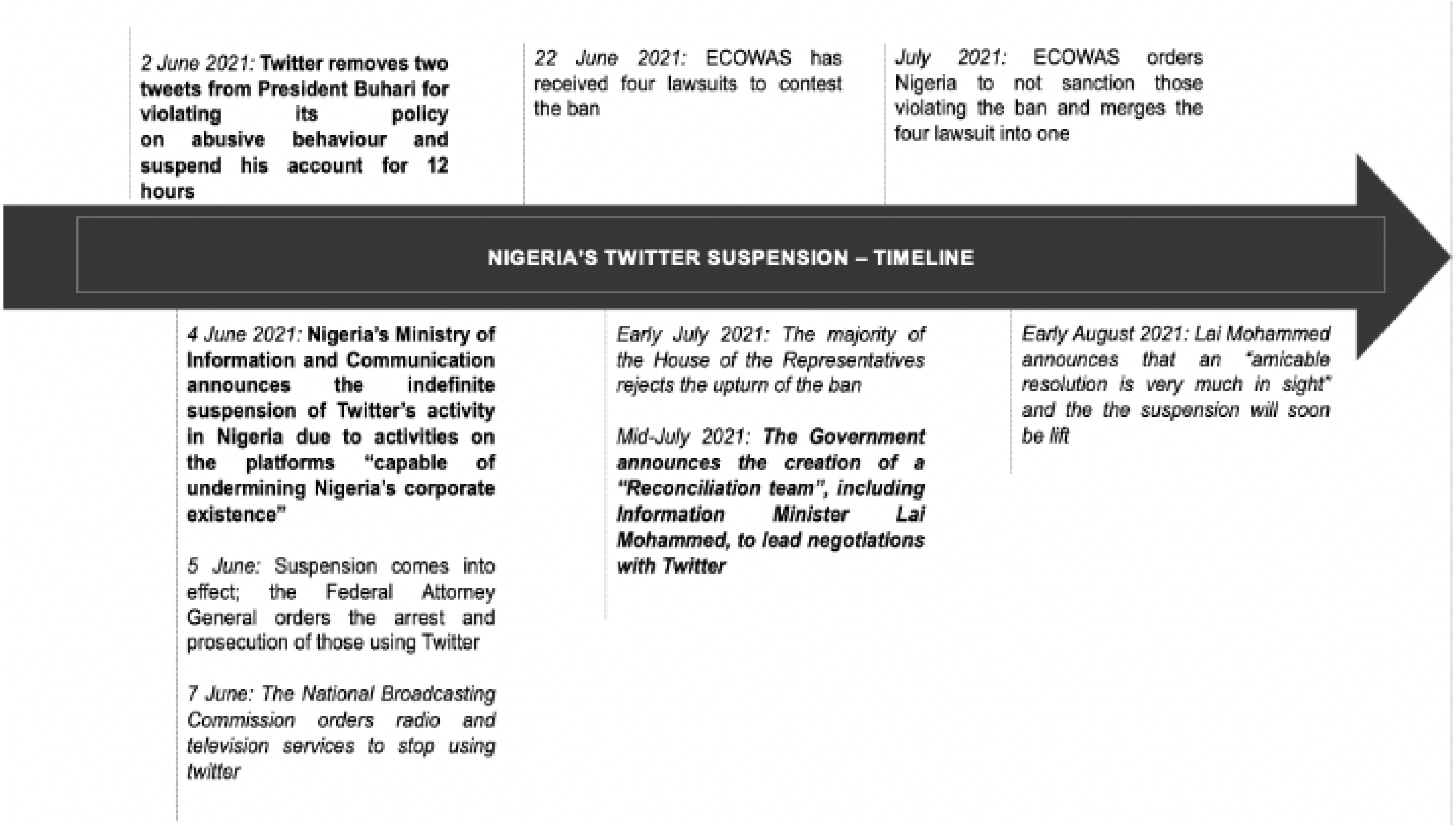
Resolution A/HRC/47/L.22 - “The promotion, protection and enjoyment of human rights on the Internet, sponsored by Nigeria

In July 2021, the United Nations Human Rights Council adopted resolution A/HRC/47/L.22 on “The promotion, protection and enjoyment of human rights on the Internet”.[379] This resolution calls for offline rights to be upheld online and consolidates related international human rights standards with a focus on internet access and shutdowns.[380] In relation to the latest events, the resolution *strongly condemns* internet shutdowns and calls for governments to refrain from using internet shutdowns and other measures to restrict access to information online. The resolution was led by a core group of Member States including Nigeria, which, at the time of its adoption, had already suspended access to Twitter.

378 https://www.article19.org/resources/nigeria-authorities-must-stop-clamping-down-on-digital-rights/

379 https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/47/L.22

380https://www.article19.org/resources/un-human-rights-council-adopts-resolution-on-human-rights-on-the-internet/



NIGERIA'S TWITTER SUSPENSION – TIMELINE

2 June 2021: **Twitter removes two tweets from President Buhari for violating its policy on abusive behaviour and suspend his account for 12 hours**

4 June 2021: **Nigeria's Ministry of Information and Communication announces the indefinite suspension of Twitter's activity in Nigeria due to activities on the platforms "capable of undermining Nigeria's corporate existence"**

5 June: Suspension comes into effect; the Federal Attorney General orders the arrest and prosecution of those using Twitter

7 June: The National Broadcasting Commission orders radio and television services to stop using twitter

22 June 2021: ECOWAS has received four lawsuits to contest the ban

Early July 2021: The majority of the House of the Representatives rejects the upturn of the ban

Mid-July 2021: ***The Government announces the creation of a "Reconciliation team", including Information Minister Lai Mohammed, to lead negotiations with Twitter***

July 2021: ECOWAS orders Nigeria to not sanction those violating the ban and merges the four lawsuit into one

Early August 2021: Lai Mohammed announces that an "amicable resolution is very much in sight" and the the suspension will soon be lift

Online Counterterrorism & Human Rights: Literature Review:

Annex 1. Literature Review on Ethical and Human Rights Risks in the use of Automated Tools in Content Moderation

1. No amount of "AI" in content moderation will solve filtering's prior-restraint problem: Emma Llansó, 23.04.2020.

This piece discusses how the technical realities of content filtering may be measured against the protections for freedom of expression in international human rights law.

2. Unboxing Artificial Intelligence: 10 steps to protect Human Rights: Council of Europe Commissioner for Human Rights, May 2019.

This recommendation on AI and human rights provides guidance to Member States on the ways in which the negative impact of AI systems on human rights can be prevented or mitigated, focusing on 10 key areas of action.

3. Artificial Intelligence & Human Rights: Opportunities & Risks: Filippo Raso, Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, and Levin Kim. Berkman Klein Center for Internet & Society at Harvard University. 25.09.2018.

This report explores the human rights impacts of AI technologies. It highlights the risks that AI, algorithms, machine learning, and related technologies may pose to human rights, while also recognizing the opportunities these technologies present to enhance the enjoyment of the rights enshrined in the Universal Declaration of Human Rights. The report draws heavily on the United Nations Guiding Principles on Business and Human Rights ("Guiding Principles") to propose a framework for identifying, mitigating, and remedying the human rights risks posed by AI.

4. Human Rights in the Age of Artificial Intelligence: AccessNow, November 2018.

AccessNow conducted this preliminary study to determine the range of AI and human rights issues that may occur today or in the near future.

5. Exploring the Human Rights Dimensions of Artificial Intelligence and Online Content Moderation at the IGF: Miru Lee, Association for Progressive Communications, 10.01.2020.

Discusses one of the agendas of the 14th annual meeting of the Internet Governance Forum (IGF): AI and human rights. According to the article, the threat to human rights and privacy because of AI was one of the main themes at the IGF. In particular, many panels discussed AI ethics and principles to protect human rights.

6. Use of AI in Online Content Moderation: Cambridge Consultants on behalf of Ofcom, 2019.

This report examines the capability of AI technologies to meet the challenges of moderating online content and how improvements are likely to enhance such capability over the next five years.

7. The impact of algorithms for online content filtering or moderation: European Parliament Policy Department for Citizens' Rights and Constitutional Affairs Directorate-General for Internal Policies, September 2020.

This study, commissioned at the request of the JURI Committee, addresses the automated filtering of online content. The report introduces automated filtering as an aspect of moderating user-generated material and presents the filtering technologies that are currently deployed to address different kinds of media, such as text, images, or videos. It discusses the main critical issues under the present legal framework and makes proposals for regulation in the context of a future EU Digital Services Act.

8. Contesting algorithms: Restoring the public interest in content filtering by artificial intelligence: Niva Elkin-Koren, 29.07.2020.

This paper discusses content moderation by AI while mentioning the hashing techniques used by GIFCT and Tech Against Terrorism. It then analyses how using AI systems to control speech raises serious concerns from a social welfare perspective.

9. Artificial Intelligence, Content Moderation, and Freedom of Expression: Emma Llansó, Joris van Hoboken, Paddy Leerssen, Jaron Harambam. Transatlantic Working Group. 26.02.2020.

This report focuses on content moderation and the use of automated systems for detecting and evaluating content at scale. It raises questions about the role of recommendation algorithms in amplifying hate speech, violent extremism, and disinformation, and further explores the use of AI and other forms of automation for both content moderation and content curation. The authors highlight issues of AI tools in connection with the risk to freedom of expression.

10. Facebook's Most Recent Transparency Report Demonstrates the Pitfalls of Automated Content Moderation: Svea Windwehr, Jillian C. York. EFF. 08.10.2020.

This paper discusses the risks posed by automated content moderation to freedom of expression online, with particular emphasis on Facebook and Instagram.

11. The Rise of Content Cartels: Evelyn Douek, 07.05.2020.

This paper traces the origin and spread of content cartels. It examines the impulses behind demands for greater cooperation and the ways in which such cooperation can be beneficial. It further explores the failures of the current arrangements and the threats they pose to free speech. GIFCT's hash-sharing database is mentioned.

12. Algorithmic content moderation: Technical and political challenges in the automation of platform governance: Robert Gorwa, Reuben Binns, Christian Katzenbach, 28.02.2020.

This article provides a technical primer on algorithmic moderation, examines some of the existing automated tools used by major platforms to handle copyright infringement, terrorism, and toxic speech, and identifies the main political and ethical issues raised by these systems as reliance on them grows.

13. Automated Moderation Must be Temporary, Transparent and Easily Appealable: Jillian C. York, Corynne McSherry. EFF. 02.04.2020.

The article recognizes that automated technology does not work at scale as it struggles to read nuance in speech the way humans can (and for some languages it barely works at all). It further notes that the use of automation results in numerous wrongful removals. The article stresses how automated moderation must therefore be temporary, transparent, and easily appealable.

14. The Limitations of Automated Tools in Content Moderation: New America.

This instalment of New America's "Everything in Moderation" series provides a more detailed discussion of the limitations of automated tools used for content moderation.

15. Promoting Fairness, Accountability, and Transparency Around Automated Content Moderation Practices: New America

In this instalment of "Everything in Moderation," New America provides a set of recommendations for developers, policymakers, and researchers to consider in pursuit of greater fairness, accountability, and transparency around algorithmic decision-making in the field of content moderation.

Annex 2. Literature Review of Ethical and Human Rights Risks in the use of Automated Tools in Content Moderation related to T/VE and Counterterrorism

1. Caught in the Net: The Impact of "Extremist" Speech Regulations on Human Rights Content: Abdul Rahman Al Jaloud, Hadi Al Khatib, Jeff Deutch, Dia Kayyali, and Jillian C. York, EFF (A joint publication from the Electronic Frontier Foundation, Syrian Archive, and Witness), 30.05.2019.

The report discusses how the reality of faulty content moderation must be addressed in ongoing efforts to address extremist content. It provides examples of blunt measures affecting marginalized users.

2. One Database to Rule them All: Svea Windwehr, Jillian C York, VOX-POL, 04.11.2020.

This article outlines concerns about GIFCT's harsh-sharing database. The concerns include reliance on automated solutions to moderate content which lead to incorrectly removing legal speech.

3. Erasing History: YouTube's Deletion of Syria War Videos Concerns Human Rights Groups

This piece discusses how thousands of videos, some of which offer crucial evidence of war crimes, have been deleted via YouTube's algorithms. In particular, it examines the hundreds of thousand videos of Syrian war atrocities that were removed by YouTube.

4. Civil Society Letter to European Parliament on Terrorism Database: AccessNow, 07.02.2019.

This open letter, from civil society organizations to the European Parliament, criticizes (in the context of the Terrorist Content Regulation debate) the blind faith in a database to flag "terrorist content." Among the concerns raised is the fact that filters are unable to understand the context and therefore are error-prone. The letter also notes the pervasive online monitoring of disadvantaged and marginalized individuals.

5. Joint Letter to EU Parliament: Vote Against Proposed Terrorist Content Online Regulation: Human Rights Watch, 25.03.2021.

In this letter to the EU Parliament, the limitations of automated content moderation tools in relation to terrorist content online are discussed.

6. Global Internet Forum to Counter Terrorism Transparency Report Raises More Questions Than Answers: Angel Diaz, Brennan Center, 25.09.2019.

This article assesses GIFCT's first transparency report and discusses the concerns about the negative impact of its hash-sharing database poses on freedom of expression.

7. The flaws in the content moderation system: The Middle East case study: Eliza Campbell, Spandana Singh, Middle East Institute, 17.11.2020.

This paper discusses the limitations of content moderation by automated tools. It emphasises that, in moderating categories of content with more fluid delineations (such as extremist propaganda and hate speech), it is difficult to develop tools that can detect or remove this content with accuracy. It examines how automated tools for content moderation impact social media users in the Middle East in particular.

8. YouTube AI deletes war crime videos as 'extremist material': Alex MacDonald, Middle East Eye, 13.08.2017.

This article discusses the criticism faced by YouTube after a new AI program monitoring "extremist" content began flagging and removing masses of videos and blocking channels that document war crimes in the Middle East.

9. Artificial Intelligence and Countering Violent Extremism: A Primer: Marie Schroeter, GNET, October 2020.

This report analyses the ability of AI applications to contribute to countering radicalization. Mapping the possibilities and limitations of this technology in its various forms, the report aims to support decision makers and experts to navigate the furore of debate, and to make informed decisions unswayed by the current hype.

Dossier F: Removing Terrorist Content

Executive Summary

Content Removal

Tech Against Terrorism developed the Terrorist Content Analytics Platform (TCAP) which alerts tech companies to terrorist content found on their platforms, and thereby supports the removal of terrorist material on the internet.

The TCAP interferes with terrorists' dissemination of terrorist content on multiple levels. Through alerting beacons, content stores, aggregators, and circumventors, the TCAP rigs with reactive measures the entire ecosystem of tech companies susceptible to exploitation by terrorists and prevents any further spread of harmful material.

To date, the TCAP has submitted **13,000** URLs and alerted **7,400** URLs to **59** tech companies. **95%** of this content has now been taken offline.

The TCAP is informed by the fundamental principles of the rule of law, transparency, privacy, and tech platform autonomy at every stage of the development process.

Tech Against Terrorism has developed comprehensive policies and protocols in line with these fundamental principles to ensure that TCAP observes the right to freedom of speech online and other ethical standards without compromising its efficacy. These measures include a group inclusion policy, content verification policy, and a threat to life protocol, all of which are underpinned by processes of extensive legal review and public consultation.

Terrorist content

Terrorism is a contested term with no universally accepted definition. There is significant academic disagreement on how to define terrorism, as it is often highly politicised.

When terrorism is defined by nation states it is often unclear what constitutes terrorist content. This makes it difficult for tech companies to moderate terrorist content and puts the onus of adjudication on tech companies.

Identifying terrorist content is not always straightforward. In addition to specific subject matter knowledge, the specific issues of context, nuance, and language make it difficult for tech platforms to make accurate assessments. This challenge is particularly difficult for smaller tech companies who do not have adequate resources to prevent terrorist abuse of their platforms.

Improvements in content moderation by tech platforms has forced terrorists and violent extremists to be highly adaptive and resilient; an array of tactics and behaviours has evolved to circumvent deplatforming, many of which enable terrorist groups to exploit the above complexities to their advantage.

RECOMMENDATIONS

Tech Against Terrorism makes the following recommendations to governments and tech companies.

Governments

1. Provide clear definitions of terrorism that are operational for tech companies to use. Ensure that key terms, such as terrorist content, are clear, practical and have a basis in existing legal frameworks, and that such definitions do not have an adverse impact on human rights and fundamental freedoms.

2. Improve and expand designation lists of proscribed terrorist organisations and provide clarity on how terrorist content produced by a designated terrorist organisation corresponds to the

online regulation in a particular jurisdiction. It can be unclear how tech companies should respond to content by any given group, which risks terrorist content staying online for longer. Tech Against Terrorism also recommends that governments consult the Consolidated United Nations Security Council Sanctions list, as it provides the best international consensus framework on terrorist groups.

3. Anchor regulation in the rule of law and ensure that it does not promote removal of legal speech via extra-legal means. Clear definitions, designation lists and other existing legal instruments will allow governments to mitigate against such risk.
4. Acknowledge the size of tech companies and the importance of proportionality when drafting requirements for tech companies. When legislation fails to do this, smaller tech companies will simply not be able to meet the requirements of legislation, which risks both terrorist content remaining online and putting smaller tech companies out of business, thereby undermining a free internet.
5. Ensure that guidance complies with appropriate safeguards of human rights and fundamental freedoms.

Tech companies

1. Introduce policies that clearly prohibit terrorism and/or violent extremism, including:
    - se of the platform by terrorist organisations and/or members
    - Expression of support for terrorist activities and/or those engaged in terrorism (material or otherwise)
    - Praise of terrorist activities and/or those engaged in terrorism
    - Promotion of terrorist activities and/or those engaged in terrorism (promotion includes posting content produced by terrorist entities)
    - Recruitment for terrorist organisations and/or terrorist activities
    - Engaging in or threatening or inciting acts of terrorism on behalf of a terrorist organization on the platform
2. Provide the company's definition of "terrorism" or "terrorist entities". Publishing a definition of terrorism, a company accountable for its content removal decisions and provides a means by which to appeal these decisions. Whilst companies may attempt to draft their own definition of terrorism, this is not a requirement. In devising an operational definition of terrorism, companies could refer to official designation lists, such as that of the United Nations Security Council and/or national designation lists, or other well-established definitions.
3. Create clearly defined parameters in content standards to provide clarity when moderating content and prevent arbitrary enforcement of these standards.
4. Ground definitions of "terrorist content" in the rule of law and have this only correspond to designated terrorist entities.

The Terrorist Content Analytics Platform (TCAP)

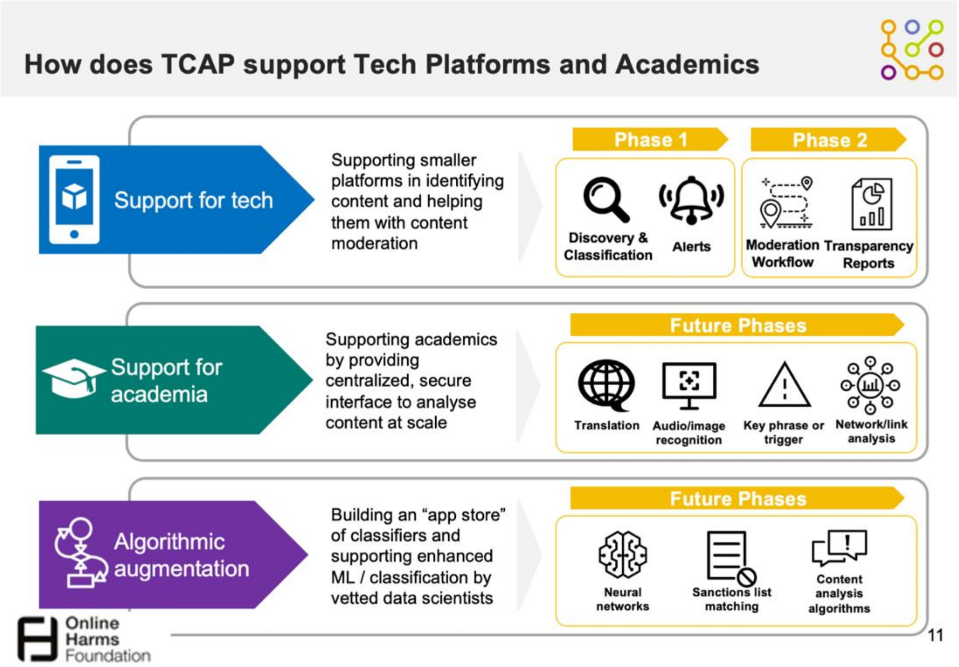
Launched in November 2020, The Terrorist Content Analytics Platform[381] (TCAP), developed by Tech Against Terrorism, assembles the world's largest database of terrorist content collected in real time from verified terrorist channels on messaging platforms and apps. As a repository of verified terrorist content (imagery, video, PDFs, URLs, audio) collected from open sources and existing datasets it facilitates secure intelligence sharing between platforms.

[381] https://www.terrorismanalytics.org/

Developed with support from Public Safety Canada,[382] TCAP is a secure online tool that automates the detection and analysis of verified terrorist content on smaller internet platforms. Following detection, the TCAP alerts tech companies to terrorist content found on their platforms and supports smaller platforms to improve their content moderation. TCAP will also improve academic research on terrorist content and augment efforts to use artificial intelligence (AI) and machine learning to detect terrorist content at scale.

The purpose of TCAP is fourfold:

1. To support tech companies in detecting terrorist content on their platforms by alerting them to terrorist content and helping to inform and manage company moderation procedures by reference to TCAP.
2. To facilitate affordable intelligence sharing for smaller internet platforms and help smaller tech companies to address terrorist use of their platforms expeditiously by means of an alert function.
3. To facilitate secure intelligence sharing between expert researchers and academics. By giving vetted academics and expert researchers access to the platform, this centralised dataset will improve quantitative analysis of terrorist use of the internet and inform the development of accurate countermeasures.
4. To facilitate the coordination of data-driven solutions to counter terrorist use of the internet by making content on the platform available as a training dataset for the development of automated solutions.



How does TCAP support Tech Platforms and Academics

Support for tech

Supporting smaller platforms in identifying content and helping them with content moderation

Support for academia

Supporting academics by providing centralized, secure interface to analyse content at scale

Algorithmic augmentation

Building an "app store" of classifiers and supporting enhanced ML / classification by vetted data scientists

Figure 38: This describes the different phases of the TCAP. Phase I was completed with the support of the government of Canada.

The problem

Terrorist and violent extremist use of the internet is increasingly concentrated on smaller platforms. Tech Against Terrorism's research shows that smaller platforms may face

[382] https://www.techagainstterrorism.org/2019/06/27/press-release-tech-against-terrorism-awarded-grant-by-the-government-of-canada-to-build-terrorist-content-analytics-platform/

disproportionately larger volumes of terrorist content on their sites, which they struggle to action due to limited capacity, capability, and subject matter knowledge.[383] Further, Tech Against Terrorism's analysis suggests that smaller tech companies struggle with technical requirements in moderating terrorist content and the solutions that are available to them.[384] Given that terrorist content will remain accessible if just one smaller tech company keeps this content online, Tech Against Terrorism concludes that all smaller tech companies need to be supported in order to counter terrorist use of the internet effectively. In response, Tech Against Terrorism has developed the TCAP.

Tech Against Terrorism's Response: The Terrorist Content Analytics Platform

The TCAP identifies terrorist content on smaller tech platforms through a URL alert function: platforms receive an automated email alert referencing the URL(s) containing the terrorist content found on the platform in question. This alert is sent to the platform as soon as Tech Against Terrorism discovers it. Platforms can also login to TCAP to view and assess all terrorist content discovered as part of Tech Against Terrorism's threat monitoring to inform their moderation decisions. TCAP also provides a content moderation workflow to facilitate moderation decisions and transparency reporting.

The key features of TCAP include:

- scraping of terrorist content from the open web
- monitoring of reported URLs and content status *i.e.*, whether a page or content is still available on the open web and if not, when it was taken offline
- manual reporting of suspicious URLs and content via Tech Against Terrorism's online portal for platforms
- automated URL alerts to platforms notifying them of terrorist content hosted on their platform.

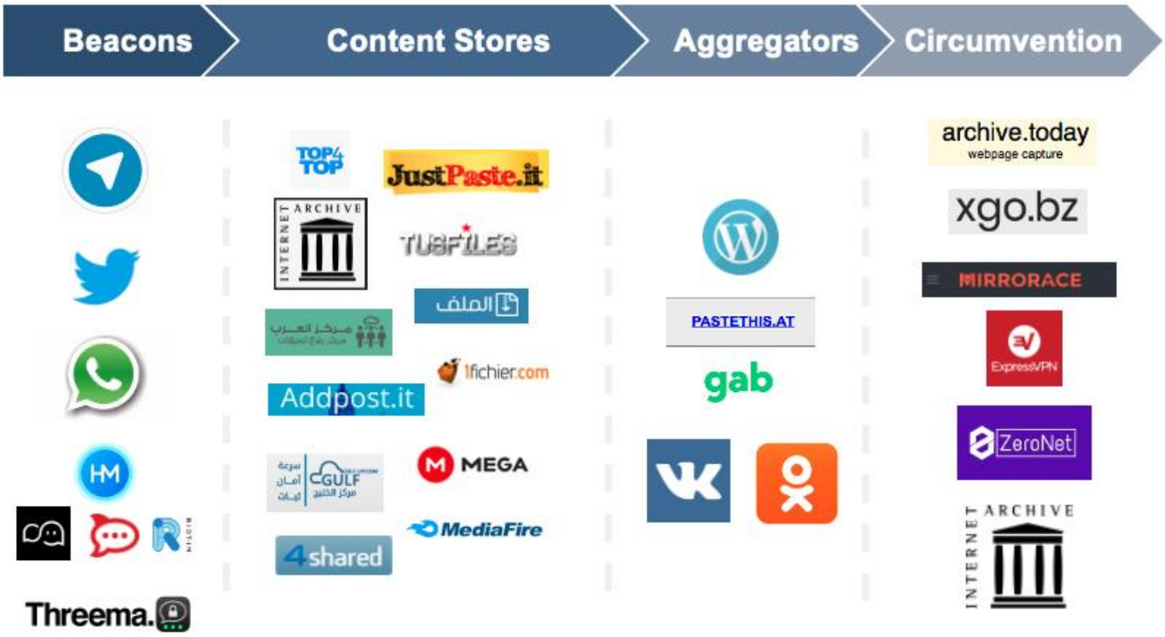
The Process: Collecting, Verifying, and Alerting Terrorist Content

Before discussing how the TCAP counters the dissemination of terrorist content, it is important to emphasise how terrorists spread their propaganda.

Terrorists and violent extremists (T/VE) have different purposes for using the internet, only one of which is propaganda dissemination. For this outwards-facing purpose, T/VE groups and actors want to reach the widest audience possible. They therefore plan their use of the internet accordingly.

Terrorists and violent extremists have long utilised the internet to spread their message, and to communicate externally. Terrorists occupy a complex and wide-reaching online ecosystem, and their presence now spans a broad range of platforms. Tech Against Terrorism has adopted the taxonomy of Dr Ali Fisher and Nico Prucha who identify three umbrella categories of tech platforms exploited by terrorists and violent extremists, to which Tech Against Terrorism has added a fourth.

[383] https://www.techagainstterrorism.org/2019/04/29/analysis-isis-use-of-smaller-platforms-and-the-dweb-to-share-terrorist-content-april-2019/
[384] https://gifct.org/wp-content/uploads/2021/07/GIFCT-TAWG-2021.pdf



Taxonomy inspired in part by A Fisher & N Prucha

Figure 39: Taxonomy of types of platforms that are used by T/VE actors

Beacons are used by terrorist and violent extremist actors to reach the widest audience possible when they disseminate their propaganda, of which WhatsApp is an example. It acts as a centrally located lighthouse and signposts to where content may be found, which is often on the content stores.

Content stores are used by terrorists and violent extremists to store their material, which includes different file types such as audio, text, images, and videos. Terrorist and violent extremists rely heavily on these platforms as this is where viewers, especially new recruits, interact and engage with terrorist material.

Aggregators act as a centralised database of where content can be found online, gathering a wide range of outlinks (URLs) to content hosting platforms to facilitate diffusion. This means that when one content store takes such content offline, it is very easy to find another platform on which this content is still online and accessible.

Circumventors complicate the potential for moderation of terrorist and violent extremist content, as, when material is taken off beacons, content stores, and aggregators, the material will often still be online and accessible on a circumventor platform. Prominent examples include the use of a VPN to access material that might be geo-blocked (only made inaccessible in a particular region), archived, or uploaded to a mirroring platform which generates outlinks to a multitude of content stores making the content more difficult to remove.

The TCAP interferes with the dissemination of terrorist content on multiple levels. First, Tech Against Terrorism's team traces terrorist groups to their preferred beacon platform, on which terrorists disseminate outlinks that direct users to smaller content stores, terrorist operated websites, and social media platforms on which the content is hosted. These platforms will then subsequently post outlinks as well, therefore the net platforms which host terrorist content grows exponentially. Through the TCAP alerts, these URLs will go offline on multiple levels, and therefore the terrorist content will be harder to find. TCAP therefore disrupts the entire ecosystem of tech platforms that terrorists use to disseminate their propaganda.

The below visualisation shows TCAP's process:

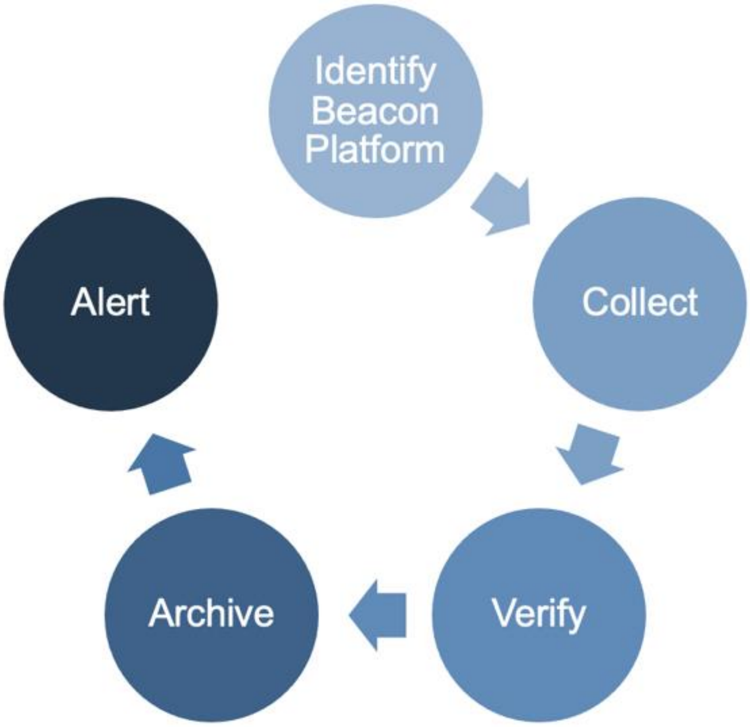


Figure 40: The TCAP's process of identifying, collecting, verifying, archiving, and alerting terrorist material

Phase I is ongoing. Tech Against Terrorism launched the TCAP automated email alerts in November 2020 and the actual platform in January in 2021. Tech Against Terrorism's future phases, support for academia and algorithmic augmentation will commence with renewed funding this autumn.

Development considerations

Consultation process

Before commencing development of TCAP, Tech Against Terrorism opened a public consultation process where tech companies, academics and members of civil society could provide feedback on what Tech Against Terrorism would need to consider when building TCAP. Questions included the scope of TCAP and what type of tools would be most useful and solicited feedback on the fundamental principles.

In August 2020, Tech Against Terrorism published a report detailing the findings from this process as part of its commitment to ensuring that the platform is developed both transparently and in full observance of human rights and fundamental freedoms, including freedom of speech.

Key findings and observations included:

Researchers and tech companies stressed that TCAP should feature tools to facilitate analysis of terrorist content, in addition to an archive of terrorist content

Researchers emphasised the need to include content spanning multiple ideologies, with a particular focus on the global violent far-right.

TCAP should be as transparent as possible, and the platform should remain independent. Respondents also underlined the importance of respecting tech platform autonomy with regard to moderation policy and enforcement decisions. As such, Tech Against Terrorism's alerts are given on an advisory basis only.

Respondents from every sector stressed the importance of safeguarding the mental health and welfare of researchers and content moderators.

Key principles in developing TCAP

Principle	Justification	Implementation
Rule of Law	The rule of law provides democratic accountability and protects fundamental human rights. As TCAP helps tech companies take content offline, it is essential that it is grounded in the rule of law to preserve these freedoms. Without this grounding, TCAP risks establishing parallel and undemocratic norms in online speech.	● Tech Against Terrorism's group inclusion policy is based on designation lists of democratic nation states and supranational organisations' designation lists - this provides tech companies with the legal grounding to remove terrorist content off their platforms and protects freedom of expression. ● We only include **official** content, using Tech Against Terrorism's content classification and verification policy.
Transparency	We want to ensure that TCAP can be held accountable for the role it plays in countering terrorist use of the internet, which we can only do through transparency. We want to ensure that stakeholders have insight into TCAP and the policies that guide it, as well as be able to give feedback on this process.[385]	● We are developing TCAP through "transparency-by-design", ensuring we are transparent in all phases of the process ● All platform policies are available on request ● We launched a public consultation process, the findings of which can be found in Tech Against Terrorism's report ● We hold monthly office hours in which we provide an update on the development of TCAP and stakeholders can ask questions and provide feedback ● Anyone with TCAP access can share their views on classification. They can contest whether a generated alert is terrorist in nature. ● At the time of writing, we are developing a transparency report which will cover the first 6 months of TCAP.
Accuracy and Accountability	To prevent undue norm setting of speech, with its inherent risks to human rights, especially freedom of expression, accuracy and accountability are vital for Tech Against Terrorism's work. We are also aware that civil society groups have cautioned that a reliance on automated	● We only raise alerts for verified content from targeted groups. These alerts come in the form of URL-sharing, importantly not hashes, so the tech company in question can review the actual content. ● We implement a vigorous verification process using in-house terrorism experts to verify the content as terrorist in

[385] At Tech Against Terrorism, we advise governments and tech companies to conduct regular transparency reports, to substantiate their transparency processes. We have launched our Transparency guidelines which considers how entities can do the same.

	tools will ultimately result in the wrongful removal of content and breaches of freedom of expression.[386]	nature, more on which can be read below in Tech Against Terrorism's policy section. ● Tech companies can "dispute content" when they think an alert has been incorrectly classified, and Tech Against Terrorism's team will review such content and keep a record for Tech Against Terrorism's transparency report. ● At the time of writing, we are setting up an Academic Advisory Board which will oversee Tech Against Terrorism's alerts, archive, and appeal process. The Board will superintend the accuracy of our alerts and their compliance with our group inclusion policy and will also adjudicate any appeals made by TCAP's users. ● We include civil society at all stages of development, to ensure we mitigate risks to human rights.
Security	Given that TCAP archives content and its location, it is imperative that we build TCAP securely, so that T/VE don't gain access to the platform. We also need to ensure that T/VE do not become aware of our operations to the extent that it inhibits our mission or risks our operational security.	● Tech Against Terrorism's follows strict OpSec protocols when conducing its open-source intelligence monitoring ● Some of its policies and office hours recordings are made available upon request, following a strict vetting process to ensure hostile actors won't be granted access. ● Tech Against Terrorism's development team executes frequent penetration testing so that TCAP as a platform can resist any attack.
Privacy	Given the often-sensitive nature of our alerts and the content we archive, we believe that data ending up in the wrong hands could lead to certain individuals being targeted by retaliatory attacks from T/VE. It is therefore critical to enforce the right to privacy.	● Any alert comes with a tag to show whether it contains personable identifiable information (PII). ● All PII will be blurred out when we release our archive later this year. ● A record of PII will be kept to preserve its potential function as digital evidence in war crimes trials or the prosecution of other human rights abuses.[387]

[386] One database to rule them all (VoxPol 2020)

[387] More on this can be read in a Human Rights Watch in a September 2020 on removal of terrorist content and war crime evidence: https://www.hrw.org/report/2020/09/10/video-unavailable/social-media-platforms-remove-evidence-war-crimes

		• PII will only be shared when we come across an immediate and credible threat to life in line with our emergency Threat to Life Protocol (see below).
Freedom of speech	We are very aware that the TCAP could pose risks freedom of expression in content moderation without sufficient safeguards in place. When tackling terrorist use of the internet it is vital that this right is respected and not undermined by extra-legal mechanisms. We aim to safeguard against "content cartels"[388] and retain the right to free expression. We are aware that we, as a non-governmental organisation, should not set global norms for online speech.	• We base our group inclusion policy on provisions of law, ensuring that we do not contribute to undue norm speech-setting of online content • We alert tech companies with the URLs containing the terrorist content so they can review the content and avoid a dependence on automated removals compromising freedom of speech. • Civil society participation ensures that relevant concerns can be raised and addressed. We ensure this participation through regular feedback sessions in office hours and our consultation report (see below) • All alerts are made on an advisory basis.
Tech platform autonomy	To avoid content 'cartelisation',[389] the TCAP alerts companies on an advisory basis only.	• All alerts are made on an advisory basis, and will explain the reason for submission as well as the relevant designation guidelines relating to the groups in question • This is supported through our Knowledge Sharing Platform[390] and Online Regulation Series[391] that makes tech platforms aware of their duties in certain jurisdictions when notified of terrorist content on their platform.

Policy considerations

Group inclusion policy

Tech Against Terrorism's group inclusion policy[392] is based on the designation lists of democratic nation states and supranational organisations. It currently includes content created by Islamic State (and official provinces), al-Qaeda (and verified affiliates), the Taliban, and designated far-right terrorist groups. The TCAP also supports the Christchurch Call to Action by notifying tech companies of material produced by the Christchurch attack perpetrator.

[388] Content cartels is a term coined by Evelyn Douek, who describes it as tech companies working together and taking content moderation decisions together without oversight. Evelyn Douek. The Rise of Content Cartels (Colombia University, 2020).

[389] See Dossier E and: https://knightcolumbia.org/content/the-rise-of-content-cartels

[390] https://ksp.techagainstterrorism.org/

[391] https://www.techagainstterrorism.org/2021/07/16/the-online-regulation-series-the-handbook/

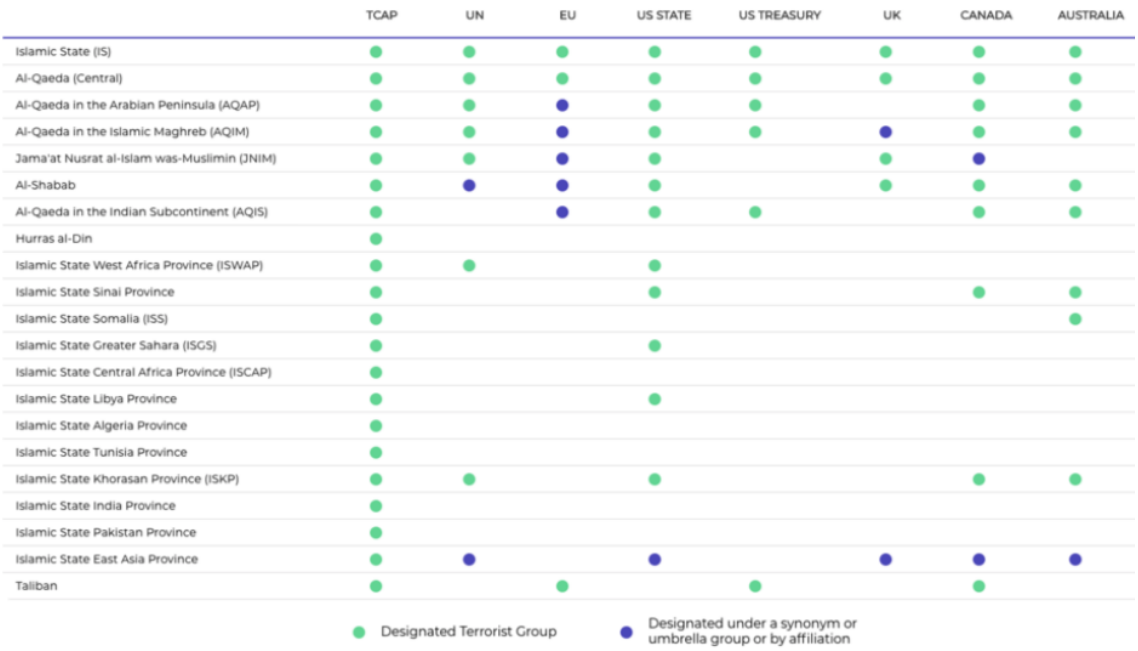
[392] https://www.terrorismanalytics.org/group-inclusion-policy

To determine which affiliates of IS and al-Qaeda to include in the initial scoping, Tech Against Terrorism has used the following methodology:

Tech Against Terrorism has examined the designation lists of democratic nation states and transnational institutions to verify firstly whether the group was proscribed as a terrorist organisation and secondly whether they were designated as an al-Qaeda or IS affiliate to establish their precise organisational proximity.

- The lists consulted include those of the United Nations, the European Union and the governments of the United States (both the State Department and Treasury), the United Kingdom, Canada, Australia and France
- Tech Against Terrorism has conducted Open-Source Intelligence (OSINT) analysis of terrorist groups' propaganda outlets and their methods of dissemination to establish official ties to al-Qaeda or IS
- Tech Against Terrorism has consulted with leading experts on terrorist groups

The following infographic shows the Islamist groups in scope and where they are currently designated.

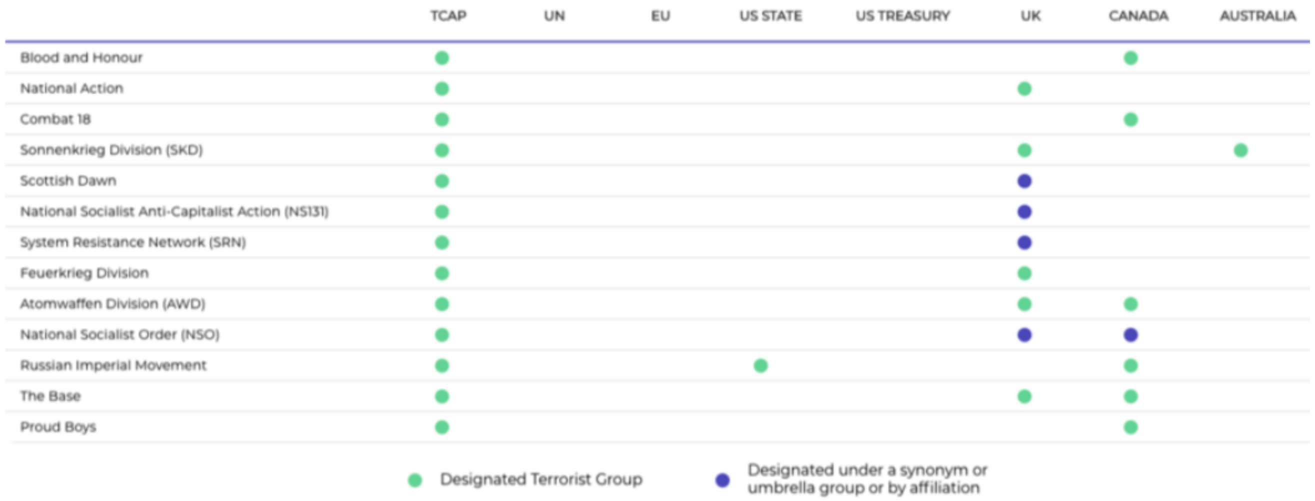


	TCAP	UN	EU	US STATE	US TREASURY	UK	CANADA	AUSTRALIA
Islamic State (IS)	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
Al-Qaeda (Central)	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
Al-Qaeda in the Arabian Peninsula (AQAP)	🟢	🟢	🔵	🟢	🟢		🟢	🟢
Al-Qaeda in the Islamic Maghreb (AQIM)	🟢	🟢	🔵	🟢	🟢	🔵	🟢	🟢
Jama'at Nusrat al-Islam was-Muslimin (JNIM)	🟢	🟢	🔵	🟢		🟢	🔵	
Al-Shabab	🟢	🔵	🔵	🟢		🟢	🟢	🟢
Al-Qaeda in the Indian Subcontinent (AQIS)	🟢		🔵	🟢	🟢		🟢	🟢
Hurras al-Din	🟢							
Islamic State West Africa Province (ISWAP)	🟢	🟢		🟢				
Islamic State Sinai Province	🟢			🟢			🟢	🟢
Islamic State Somalia (ISS)	🟢							🟢
Islamic State Greater Sahara (ISGS)	🟢			🟢				
Islamic State Central Africa Province (ISCAP)	🟢							
Islamic State Libya Province	🟢			🟢				
Islamic State Algeria Province	🟢							
Islamic State Tunisia Province	🟢							
Islamic State Khorasan Province (ISKP)	🟢	🟢		🟢			🟢	🟢
Islamic State India Province	🟢							
Islamic State Pakistan Province	🟢							
Islamic State East Asia Province	🟢	🔵		🔵		🔵	🔵	🔵
Taliban	🟢		🟢		🟢		🟢	

🟢 Designated Terrorist Group   🔵 Designated under a synonym or umbrella group or by affiliation

Figure 41: The designated Islamist terrorist organisations in scope of the TCAP.

Tech Against Terrorism has only included far-right terrorist groups that have been designated as such by a country it assesses to be a democratic country or by a supranational organisation. Tech Against Terrorism has update its group inclusion policy in accordance with evolving trends in designation. The following infographic shows the violent far-right groups that Tech Against Terrorism targets and where they are currently designated.



	TCAP	UN	EU	US STATE	US TREASURY	UK	CANADA	AUSTRALIA
Blood and Honour	🟢						🟢	
National Action	🟢					🟢		
Combat 18	🟢						🟢	
Sonnenkrieg Division (SKD)	🟢					🟢		🟢
Scottish Dawn	🟢					🔵		
National Socialist Anti-Capitalist Action (NS131)	🟢					🔵		
System Resistance Network (SRN)	🟢					🔵		
Feuerkrieg Division	🟢					🟢		
Atomwaffen Division (AWD)	🟢					🟢	🟢	
National Socialist Order (NSO)	🟢					🔵	🔵	
Russian Imperial Movement	🟢			🟢			🟢	
The Base	🟢					🟢	🟢	
Proud Boys	🟢						🟢	

🟢 Designated Terrorist Group 🔵 Designated under a synonym or umbrella group or by affiliation

Figure 42: The designated far-right terrorist organisations in scope of the TCAP.

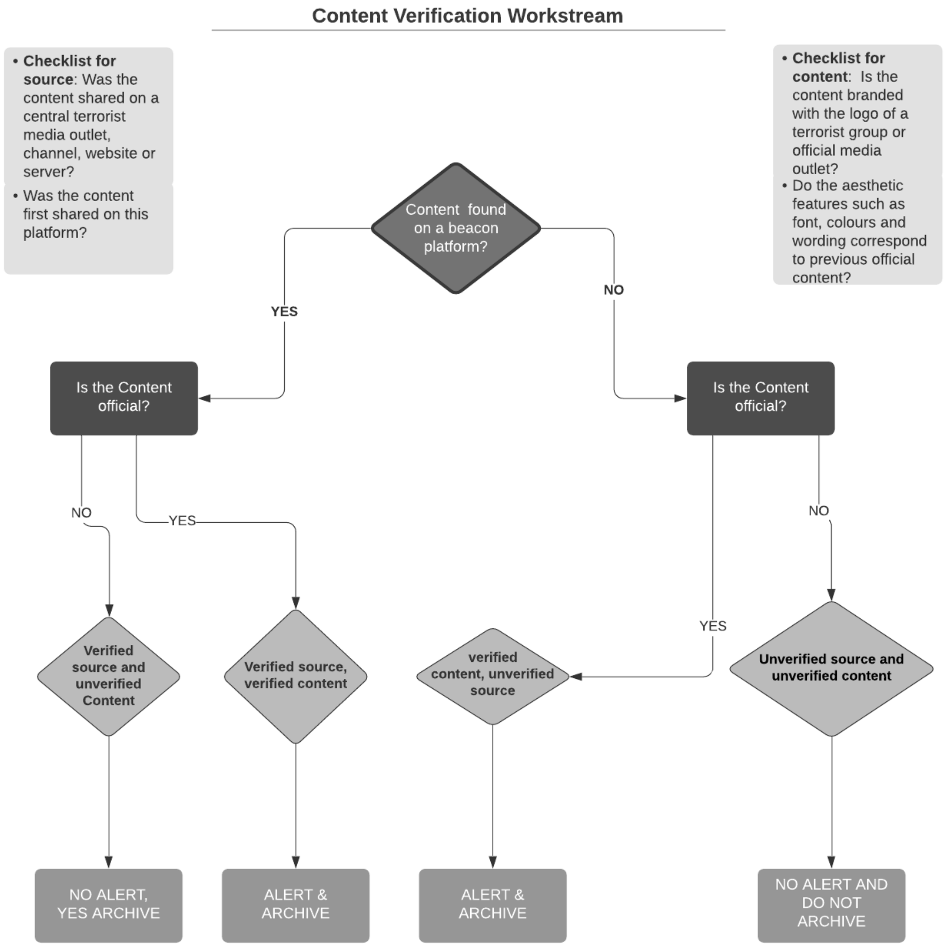
Tech Against Terrorism may in future expand its policy to include:

- Additional ideological strands
- weaker affiliations to designated far-right and Islamist terrorist groups
- supporter-generated material rather than just official material produced by a targeted terrorist group
- online material produced by verified lone-actor terrorists including manifestos and livestreams.

Content classification and verification policy

Tech Against Terrorism's content classification and verification policies operate in tandem with the group inclusion policy to ensure that only official content is submitted to TCAP. Official content is the material produced by a terrorist group or their media agency and differs from supporter-generated material, which is material published in support of a terrorist organisation. Tech Against Terrorism's content classification and verification policy guides the analysis of content in the TCAP.

This process is reflected below:



Content Verification Workstream
• Checklist for source: Was the content shared on a central terrorist media outlet, channel, website or server?
• Was the content first shared on this platform?
• Checklist for content: Is the content branded with the logo of a terrorist group or official media outlet?
• Do the aesthetic features such as font, colours and wording correspond to previous official content?
Content found on a beacon platform?
YES
NO
Is the Content official?
Is the Content official?
NO
YES
YES
NO
Verified source and unverified Content
Verified source, verified content
verified content, unverified source
Unverified source and unverified content
NO ALERT, YES ARCHIVE
ALERT & ARCHIVE
ALERT & ARCHIVE
NO ALERT AND DO NOT ARCHIVE

Figure 43: Content Verification Workstream

Threat to life protocol

Tech Against Terrorism has a protocol to guide its Open-Source intelligence team on what to do when they encounter a potential threat to life.

A threat to life is determined as a deliberate intention to cause:

- A real and immediate threat to life (real and immediate defined as a risk that is reasonably assessed to be real, and the potential assailant has the intention and capability to carry out the threat)
- Threat to cause serious harm
- Threat of injury
- Threat of serious sexual assault and/or rape

Tech Against Terrorism identifies a threat to life by considering:

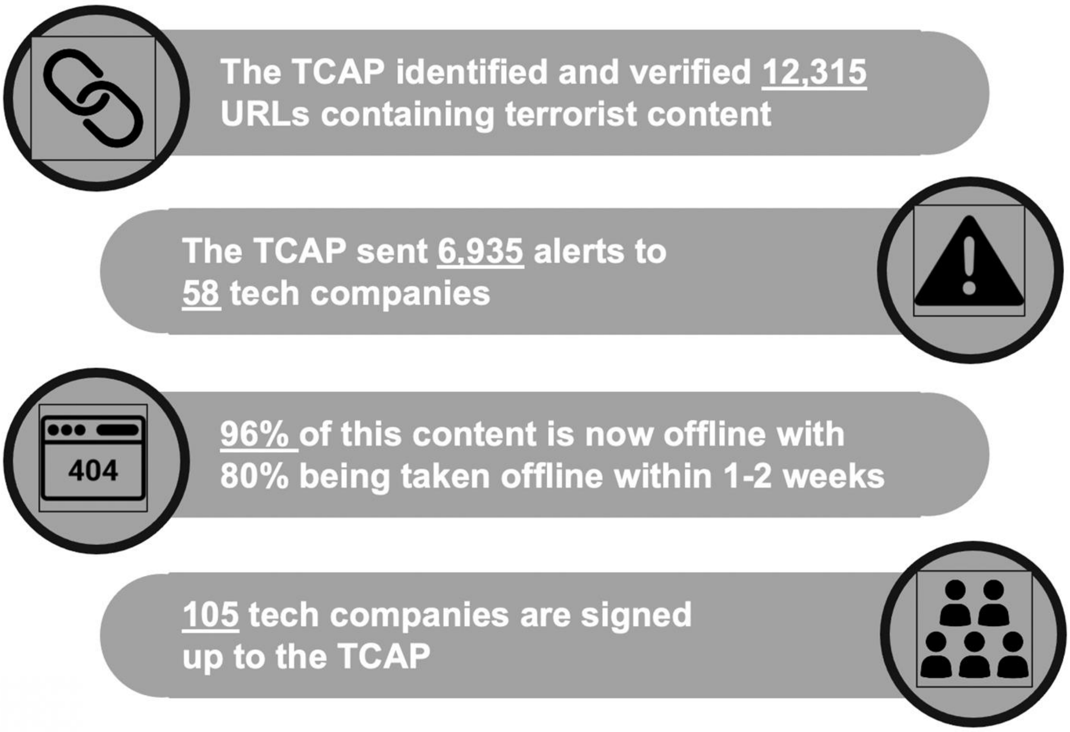
- The potential attacker and his/her capability
- The intent of the attacker
- The victim(s)
- The location

- The timescale
- Gaps in available information.
- Risk levels:
    - **Low:** no “real and immediate” threat, indicating the perpetrator has no intention or capability to follow through a threat
    - **Medium:** the alleged threat is likely to occur if the perpetrator has the right resources; the threat is therefore **conditional**. The intention will need to be assessed as “highly certain” to justify raising the threat level.
    - **High:** The threat is credible, immediate, and specific. Suspect, victim, and location of the threat is identifiable. However, there can be a high threat to life but without all this information – leading to an unspecific high threat to life.

When the TCAP team monitoring terrorist channels and content encounters a High Threat to Life, Tech Against Terrorism (which is a UK based organisation) will alert United Kingdom police and relevant authorities if the location is known.

TCAP: impact to date

Since the November 2020 launch:



The TCAP identified and verified 12,315 URLs containing terrorist content

The TCAP sent 6,935 alerts to 58 tech companies

96% of this content is now offline with 80% being taken offline within 1-2 weeks

105 tech companies are signed up to the TCAP

Statistics per month:

Month	Total Submissions	Total Alerts	% Offline
January	1,138	526	95%
February	2,104	1,036	83%
March	1,967	806	91.5%
April	1,737	1,044	90%

May	1,831	905	90%
June	1,234	811	81%
July	1,875	1,003	83%
August	1,606	929	90%

TCAP has also received significant recognition from stakeholders.

- TCAP was commended by Prime Minister of Canada Justin Trudeau at the Christchurch Call to Action 2021 Summit.
- TCAP was mentioned in a report by Human Rights Watch, which states that the considerations taken by Tech Against Terrorism in building the TCAP will be informative for Human Rights Watch' mechanism to archive removed material for evidence of war crimes.
- TCAP was also mentioned in the Digital Lockers Human Rights Report which discusses 'Voluntary Partnership Models' in archiving media evidence of 'Atrocity Crimes'. The report was published by UC Berkeley, and is accessible here.
- TCAP's contribution to countering Islamic State's propaganda was highlighted by the United Nations Counter-Terrorism Directorate (UNCTED) in their Twelfth report on the Threat posed by ISIL, Daesh to international peace and security.

Terrorist content: definitions and identification challenges

What is terrorist content?

Identifying what is or isn't terrorist content is often tied to how terrorism itself is defined. Terrorism is a contested term without a universally accepted definition. There is significant academic disagreement on how to define terrorism, as it is often highly politicised.

Many governments provide definitions of terrorism in national legislation, but it is often unclear how content can be classified accordingly. This leaves the onus of adjudicating what constitutes terrorist content on tech companies, which Tech Against Terrorism views as an abdication of governmental responsibility. The content moderation challenges faced by tech companies will be explored in the following section.

When countries do have a definition of terrorist content, they are sometimes impractically broad and circular. This presents serious risks for freedom of expression as these definitions could be used to pressure tech companies to remove legal or non-violent speech. Some nations do not have a definition of 'terrorist content' in their legislation, and refer instead to illegal, offensive, harmful, or violent abhorrent materials. These broad terms will often refer to violent content but do not contain any provisions for non-violent terrorist content. In nations that focus purely on violent terrorist content, much of the 'grey area' content may not be included under the rule.[393]

As an example of the many definitions adopted by governments, the below table shows a selection of different terms that countries and inter-governmental organisations use to define terrorism or terrorist content.

[393] For an in-depth discussion of this, see Dossier B and Tech Against Terrorism's Online Regulation Series Handbook: https://www.techagainstterrorism.org/wp-content/uploads/2021/07/Tech-Against-Terrorism-%E2%80%93-The-Online-Regulation-Series-%E2%80%93-The-Handbook-2021.pdf

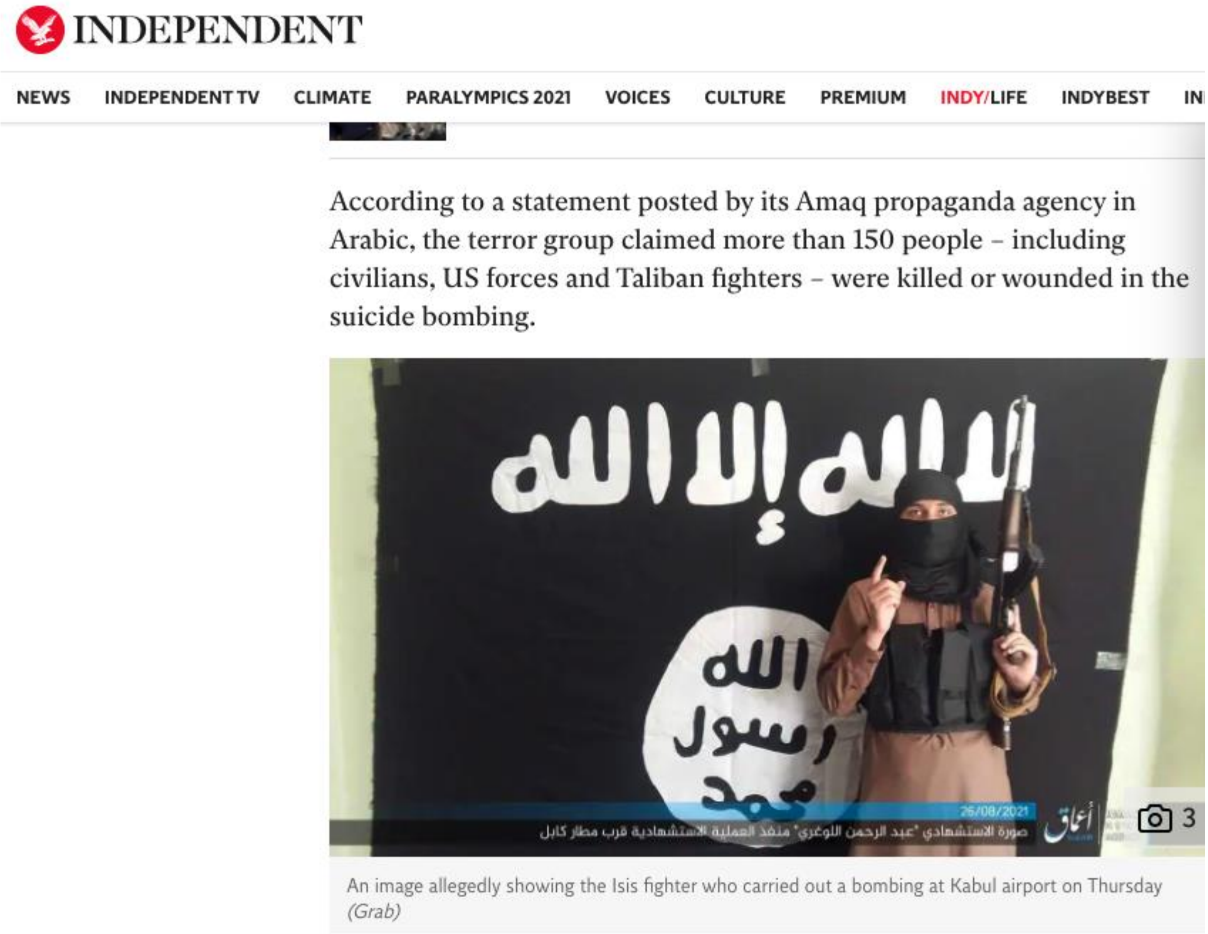
Country	Definition of terrorism or terrorism content
United Kingdom	"Illegal content is 'terrorism content' if the relevant offence is a terrorism offence."[394]
Canada	Terrorist propaganda defined as "any writing, sign, visible representation or audio recording that counsels the commission of a terrorism offence."[395]
New Zealand	"Objectionable Content" relating to harmful or graphic content.[396] Weight is given to whether the content "promotes or encourages criminal acts or acts of terrorism."
European Union	Terrorist content means material that "directly or indirectly, such as by the glorification of terrorist acts, advocates the commission of terrorism offences."[397]

Identification challenges: nuance, context, and accuracy

Terrorist content has context which will need to be considered when tech companies make content moderation decisions. The problem is that smaller tech companies often don't have the capacity to comprehend such context.

An example of the importance of context is the sharing of terrorist material for academic or journalistic purposes. In such instances it may be difficult for companies to assess whether content is terrorist propaganda, or whether it is part of journalistic reporting or academic research on terrorism, or civil society efforts to collect evidence of war crimes and human rights abuses. When platforms fail to make this distinction, they are often criticised; however, there are as of yet no clear guidelines to assist platforms on how to make such distinctions, particularly when their audiences are international.

[394] Online Safety Bill , United Kingdom, 2021, p.38
[395] Criminal Code, Canada, 2019, 144
[396] Films, videos, and publications Act, New Zealand, 1983 (3).
[397] TERREG: Preventing the dissemination of terrorist content online, The European Commission, 2021.



INDEPENDENT

NEWS INDEPENDENT TV CLIMATE PARALYMPICS 2021 VOICES CULTURE PREMIUM INDY/LIFE INDYBEST IN

According to a statement posted by its Amaq propaganda agency in Arabic, the terror group claimed more than 150 people – including civilians, US forces and Taliban fighters – were killed or wounded in the suicide bombing.

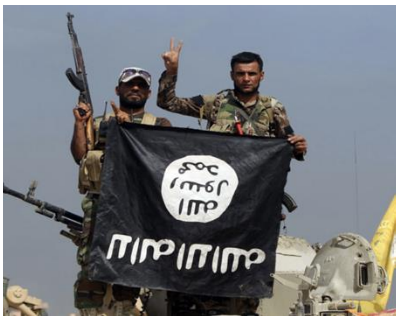
لا إله إلا الله
الله
رسول
محمد
26/08/2021
صورة الاستشهادي "عبد الرحمن اللوغري" منفذ العملية الاستشهادية قرب مطار كابل
أعماق
3

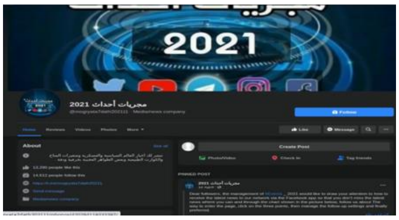
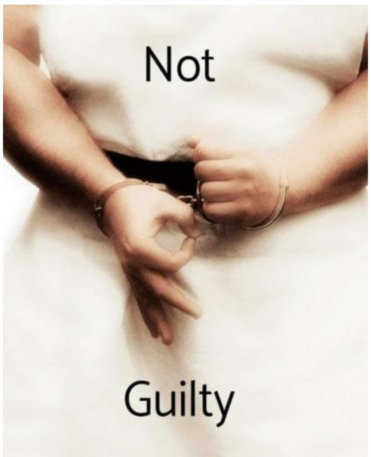
An image allegedly showing the Isis fighter who carried out a bombing at Kabul airport on Thursday *(Grab)*

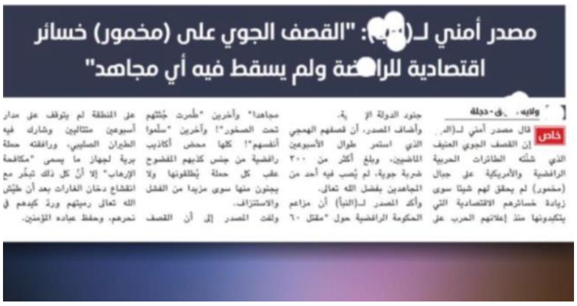
Figure 44: The Independent, a British newspaper, sharing Islamic State' material that indicates the group claimed the attack on Kabul airport in Afghanistan, on 26 August 2021.

It is important to emphasise that even if platforms provide a definition for terrorist groups, terrorist propaganda comes in many forms and each piece of content contains nuances and depends heavily on context.

Terrorists and violent extremists are also aware of this and are skilled at ensuring that their content stays within what is allowed. When having to determine what terrorist content means, some tech companies may be able to easily detect violent content, but with most terrorist groups frequently sharing "grey area" content and some non-violent material, this makes it difficult to identify. The following table provides examples of where the context of the material in question can make identification and adjudication challenging.

Identification issue	Explanation	Example
Non-violent propaganda material	Though some terrorist content depicts violence which would be prohibited under many platforms' terms of service, other forms of propaganda can go undetected as they do not directly depict violence. This is why many platforms have chosen to categorise terrorist content in varying forms such as "use", "support", "praise", "promote", "recruit", "engage", "threaten" or "incite" – all of which have been demonstrated in different ways by examples above.	An IS propaganda video detailing the caliphate's medical services. As it doesn't depict explicit violence, it can be hard for a tech company to understand the terroristic nature of this content and the propaganda value this has.
News content and/or content shared to document human rights abuses	It may be difficult for companies to assess whether content is terrorist propaganda, or whether it is part of journalistic reporting or academic research on terrorism, or civil society efforts to collect evidence of war crimes and human rights abuses.	The image below shows Kurdish militias showing the Islamic State flag upside down, which is a sign of disrespect and shows how the Kurdish militias are regaining ground that was previously ruled by IS. Tech companies may remove this content automatically, as the symbol of the IS flag may be detected by content moderation algorithms. This content has value as both journalism and potentially as an assertion of human rights and ought therefore to stay up. 

Content that appears journalistic in nature but is actually terrorist content	At Tech Against Terrorism, we have identified a new trend whereby T/VE actors pretend to be journalists and promote their content as journalistic in nature. It can be difficult for tech companies to understand that this is the case, especially in non-Western languages.	The below screenshot shows a Facebook page that was created by Islamic State supporters and pretends to be a legitimate news agency.  مجريات أحداث 2021 Dear followers, the management of [illegible] 2021 would like to draw your attention to how to receive the latest news to our network via the Facebook app so that you don't miss the latest news where you can and through the chart shown in the picture below, follow us about The way to enter the page, click on the three points, then manage the follow-up settings and finally preferred. 2021 Follow Home Reviews Videos Photos More Like Message About See all Create Post Photo/Video Check in Tag friends PINNED POST
Use of niche imagery and symbols	Terrorist groups and actors often use imagery and symbols that can be hard for a tech company moderator to identify.	The following image shows Christchurch attacker Brenton Tarrant making a hand gesture during a court appearance. The gesture shows the "OK-sign", a common symbol used by the extremist far-right indicating "white power". Seen on its own, without sufficient context it is challenging for tech companies to identify and moderate. 

Terrorist content in other languages	Whereas bigger tech companies may have the capacity and resources to invest in language recognition, smaller tech companies often do not have this capability. This makes it difficult for smaller tech companies to swiftly respond to terrorist content in other languages.	The following screenshot shows Islamic State propaganda in which the logos of their media outlet have been blurred out. This means that for a tech company to identify this as terrorist content, they would need to be able to read Arabic to examine the text.  مصدر أمني لـ( ): "القصف الجوي على (مخمور) خسائر اقتصادية للرا ضة ولم يسقط فيه أي مجاهد" ولاية ... دجلة خاص قال مصدر أمني لـ(ال... إن القصف الجوي العنيف الذي شنّه الطائرات الحربية الرافضية والأمريكية على جبال (مخمور) لم يحقق لهم شيئا سوى زيادة خسائرهم الاقتصادية التي يتكبدونها منذ إعلانهم الحرب على جنود الدولة الإ ية. وأضاف المصدر، أن قصفهم الهمجي الذي استمر طوال الأسبوعين الماضيين، وبلغ أكثر من ٣٠٠ ضربة جوية، لم يُصب فيه أحد من المجاهدين بفضل الله تعالى. وأكد المصدر لـ(النبأ) أن مزاعم الحكومة الرافضية حول "مقتل ٦٠ مجاهدا" وآخرين "طُمرت جثثهم تحت الصخور"! وآخرين "سلّموا أنفسهم"! كلها محض أكاذيب رافضية من جنس كذبهم المفضوح عقب كل حملة يُطلقونها ولا يجنون منها سوى مزيدا من الفشل والاستنزاف. ولفت المصدر إلى أن القصف على المنطقة لم يتوقف على مدار أسبوعين متتاليين وشارك فيه الطيران الصليبي، ورافقته حملة برية لجهاز ما يسمى "مكافحة الإرهاب" إلا أن كل ذلك تبخّر مع انقشاع دخان الغارات بعد أن طيّش الله تعالى رميتهم ورد كيدهم في نحرهم، وحفظ عباده المؤمنين.

Content Moderation Avoidance Techniques

As many platforms have improved their response to terrorist content, terrorists and violent extremists have developed techniques to avoid content moderation which impedes its identification by tech companies. The table below identifies the main techniques utilised by terrorists and extremists.

Tactic	Description
Mirroring	Anticipating that their accounts, channels, servers or posts are likely to be taken down by platform administrators, terrorists and violent extremists sometimes create multiple identical accounts, or upload multiple copies of the same content simultaneously. The aim is to overwhelm content moderation teams by creating more accounts than they are capable of moderating. This tactic has been pioneered by Islamic State on Telegram in particular, where it has simultaneously run as many as 20 mirror versions of its ‘official’ propaganda channel.
Private channels and/or servers	Terrorist and violent extremist organisations and groups will often respond to takedowns of public groups and channels by creating private, invite-only versions. Depending on the platform, this will make it more difficult for content moderation teams to take down the channel or group, particularly when the channel name does not provide clues to its content; some platforms do not moderate private channels at all. Share links to the channel can be shared within and outside the platform. Examples include private invite links on both Rocketchat and Telegram.
Content editing and repurposing	Content produced by terrorist organisations is often edited and repurposed to avoid automated takedowns, for example by blocking

	out branding or segmenting illegal content from that which is more admissible, such as mainstream media reporting. On Telegram, for example, public pro-IS channels often blur out the logo of Amaq News, an official IS outlet, from the top right corner of video productions.
Language amendments	Terrorists and violent extremists avoid keyword detection by tech platforms by amending terms and phrases that may already be on the radar of content moderation teams. They may insert spaces and underscores in the middle of key phrases, for example, or change their language entirely. We have seen Telegram channels containing Arabic IS content, for example, change their titles to Mandarin. Another prominent example is the ‘Boogaloo’ movement, which adapted its title to other similar iterations such as the ‘Big Luau’ and the ‘Big Igloo”.
Rhetoric dilution	Knowing the Terms of Service of the platforms on which they operate, many extremist individuals and organisations intentionally dilute their rhetoric to avoid deplatforming. This is despite their rhetoric being often more overtly hateful or supportive of violence elsewhere. This is particularly the case with far-right (violent) extremists, who attempt to pose as legitimate, non-racist political commentators on mainstream platforms while posting more extreme content elsewhere.[398]
Misrepresentation	Terrorists and violent extremists often exploit legal clauses in several countries that allow a defence of terrorist content by invoking the purposes of journalism or research. Violent far-right extremists, for example, often share graphic content or instructional material alongside a deliberate caveat that they are sharing for ‘journalistic’ purposes, and that they ‘do not endorse’ the material being shared. Among violent Islamists, the pro-al-Shabaab website Somali Memo provides an instructive example. The website presents itself as a legitimate and impartial Somali news website, but its reporting is almost exclusively based on the propaganda output of al-Shabaab’s official media outlets. It also publicises al-Shabaab propaganda videos in full. In our view this is a deliberate misrepresentation tactic intended to circumvent content moderation.
Outlinking	By posting content via third-party platform outlinks, terrorists may evade detection by content moderation teams, particularly when the linked content would be picked up by automated detection systems if it were posted

[398] Mark Collett, for example, is a British neo-Nazi activist who maintains accounts on Telegram and, until recently, Twitter. Collett’s messaging on Twitter was often sanitised, for example by only selecting mainstream media reporting that aligned with his worldview. On Telegram, on the other hand, he speaks more freely in relation to his racist and anti-Semitic views.

	in-app. As outlined above, terrorists and violent extremists also often post multiple outlinks to the same content simultaneously, in the knowledge that the content is likely to be taken down. This tactic increases the chance that the content can be found on at least one of the outlinks. See the screenshots above which show the distribution of content at different resolutions, using mirror platforms. These show the use of outlinks by terrorist groups.
Archiving	Web archiving services such as the Internet Archive are used by terrorists and violent extremists to create backed-up copies of content that has been uploaded to file-sharing platforms. Many of these services are free and easy to use, and guarantee user anonymity.

Commonwealth Secretariat
Marlborough House, Pall Mall
London SW1Y 5HX
United Kingdom

thecommonwealth.org

The Commonwealth