Global Network
on Extremism & Technology

Decoding Hate: Using Experimental Text Analysis to Classify Terrorist Content

Abdullah Alrhmoun, Shiraz Maher, Charlie Winter

GNET is a special project delivered by the International Centre for the Study of Radicalisation, King's College London.

The authors of this report are Abdullah Alrhmoun, a doctoral researcher at the Central European University in Vienna, Austria; Dr. Shiraz Maher, Director of the International Centre for the Study of Radicalisation (ICSR) at King's College London; and Dr. Charlie Winter, Senior Research Fellow at ICSR

The Global Network on Extremism and Technology (GNET) is an academic research initiative backed by the Global Internet Forum to Counter Terrorism (GIFCT), an independent but industry-funded initiative for better understanding, and counteracting, terrorist use of technology. GNET is convened and led by the International Centre for the Study of Radicalisation (ICSR), an academic research centre based within the Department of War Studies at King's College London. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing those, either expressed or implied, of GIFCT, GNET or ICSR.

This work was supported by a research award from Facebook as part of its 'Content Policy Research on Social Media Platforms' research project. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the policies, either expressed or implied, of Facebook.

CONTACT DETAILS

For questions, queries and additional copies of this report, please contact:

ICSR
King's College London
Strand
London WC2R 2LS
United Kingdom

T. **+44 20 7848 2098**
E. **mail@gnet-research.org**

Twitter: **@GNET_research**

Like all other GNET publications, this report can be downloaded free of charge from the GNET website at www.gnet-research.org.

Contents

1 Introduction

This paper uses automated text analysis – the process by which unstructured text is extracted, organised and processed into a meaningful format – to develop tools capable of analysing Islamic State (IS) propaganda at scale.[1] Although we have used a static archive of IS material, the underlying principle is that these techniques can be deployed against content produced by any number of violent extremist movements in real-time. This study therefore aims to complement work that looks at technology-driven strategies employed by social media, video-hosting and file-sharing platforms to tackle violent extremist content disseminators.[2] In general, these platforms aim to remove material produced by terrorist and hate organisations unless such material is disseminated in very specific contexts (such as, for instance, by journalists or academics).[3] In recent years, the collective efforts of such platforms have become highly effective, with almost all terrorist content being removed before it has even been reported.[4]

However, not all terrorist content is created equal.[5] The removal of certain material needs to be prioritised.[6] Problems of automation can arise, particularly in cases where technology is used to identify harmful content, which is then put before human reviewers to make a final decision; such a process can present a serious challenge around questions of what is prioritised for review, and how and when it is removed.[7] Put another way, can technology be developed to assess the material effectively and accurately? A distinction needs to be drawn between materials that need immediate review and materials that can be placed in a queue.[8] For example, one might consider the relative risk of a photograph showing a somewhat benign image of terrorist socialisation compared to a video depicting graphic violence.

Moreover, where automation has played a role, it has traditionally been deployed against the *context* in which social media posts exist rather than in relation to the *content*. As a result, these enquiries do not adequately harness the sorts of tools that are most practicable for tech company moderators.

1 Justin Grimmer and Gary King, 'General Purpose Computer-Assisted Clustering and Conceptualization', Proceedings of the National Academy of Sciences, 2011. Accessed at: https://j.mp/2nRjqbO; Gary King and Justin Grimmer, 'Method and Apparatus for Selecting Clusterings to Classify A Predetermined Data Set', United States of America 8,438,162 (May 7), 2013. Accessed at: https://j.mp/2ovzAuR.

2 For an example of this, see: James Vincent, 'UK creates machine learning algorithm for small video sites to detect ISIS propaganda', The Verge, 13 February 2018. Accessed at: https://www.theverge.com/2018/2/13/17007136/uk-government-machine-learning-algorithm-isis-propaganda.

3 Guy Rosen, 'How are we doing at enforcing our community standards?' Facebook, 15 November 2018. Accessed at: https://newsroom.fb.com/news/2018/11/enforcing-our-community-standards-2/.

4 Monica Bickert, 'Hard questions: What are we doing to stay ahead of terrorists?' Facebook, 8 November 2018. Accessed at: https://newsroom.fb.com/news/2018/11/staying-ahead-of-terrorists/.

5 See, for example: Charlie Winter, 'Understanding jihadi stratcom: The case of the Islamic State', Perspectives on Terrorism vol. 13:1 (2019): 54–62; Charlie Winter and Dounia Mahlouly, 'A tale of two caliphates: Comparing the Islamic State's internal and external messaging priorities'. VOX-Pol, July 2019; Stephane Baele and Charlie Winter, 'From music to books, from pictures to numbers: The forgotten—yet crucial—components of IS' propaganda', in Stephane Baele, Travis Coane,and Katharine Boyd (eds.), The Propaganda of the Islamic State, Oxford: Oxford University Press (2019).

6 Bickert. 'Hard questions'.

7 Alex Schulz and Guy Rosen, 'Understanding the Facebook: Community standards enforcement report', Facebook, May 2020. Accessed at: https://fbnewsroomus.files.wordpress.com/2018/05/understanding_the_community_standards_enforcement_report.pdf. p.17.

8 Bickert, 'Hard questions'.

This project explores how tech companies can draw that distinction in a timely and accurate fashion. Using IS as an initial test-case, it will add nuance to how harmful content is classified. Our basic premise is that it is possible to codify the intent of harmful content by studying and then testing the logic behind its production.[9] Indeed, if intent can be identified – that is, if a clear distinction can be drawn between tactical, action-based content and strategic, brand-based content[10] – then it will be possible to better prioritise review and removal according to risk posed.

To this end, we aim to synthesise subject-matter expertise with data science in this paper, using experimental text processing to interrogate and categorise our repository of official IS content. Our principal objective is to develop automated methods that can, when applied to similar bodies of materials (including those that are much larger), accelerate the process by which they may be disaggregated and, where relevant, triaged for moderation and/or referral. This will, in turn, help to enhance existing content moderation policies.

Given the overwhelming volume of content produced minute by minute, the need for such an approach is clearly pressing. On YouTube alone more than 300 hours of video are uploaded to the platform every minute, with users watching over a billion hours of video every day.[11] There are, on average, 500 million tweets produced per day, totalling around 200 billion per year.[12] During the first quarter of 2020, Facebook registered more than 2.6 billion monthly active users, while Instagram, which is owned by Facebook, hosts more than 500 million 'Instagram Stories' every day.[13] The overwhelming majority of users on these platforms are, of course, there for wholly benign and legitimate purposes. There is no suggestion that this content needs to be censored or monitored in any way. However, the presence of malevolent violent extremists operating within the midst of this tsunami of material demonstrates the need for effective automated methods that are able to identify, parse and disaggregate the range of content that they produce.

This is why we have chosen to focus on IS material, which brought the issue of violent extremist content into sharp relief. On the whole, the broader jihadist movement has harnessed the use of technology for propaganda purposes better than other movements.[14] In the 1990s, static websites, such as Azzam.com, brought news of jihadist campaigns in Chechnya, Bosnia and Afghanistan to English-speaking audiences. Following the 2003 invasion of Iraq,

9 As Berger notes, extremisms vary widely in terms of their respective narratives, but little in terms of the structures in which those narratives are couched. Future iterations of the classifier could be developed to assist in Facebook's content policies and practices regarding other forms of extremism. See: JM Berger, Extremism, Cambridge: The MIT Press (2018): 51–112.

10 For an introductory account of this distinction, see: Winter, 'Understanding jihadi stratcom'.

11 Data collected on 11 August 2020 at 'YouTube for Press'. Accessed at: https://www.youtube.com/intl/en-GB/about/press/.

12 David Sayce, 'The number of tweets per day in 2020', David Sayce, May 2020. Accessed at: https://www.dsayce.com/social-media/tweets-day/.

13 Jessica Clement, 'The number of monthly active Facebook users worldwide as of the first quarter of 2020', Statista, 10 August 2020. Accessed at: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/; Maryam Mohsin, '10 Instagram Stats Every Marketer Should Know in 2020', Oberlo, 6 February 2020. Accessed at: https://www.oberlo.co.uk/blog/instagram-stats-every-marketer-should-know.

14 On the tech side of the equation, see: Brian Fishman, 'Crossroads: Counter-terrorism and the Internet', Texas National Security Review, February 2019. Accessed at: https://tnsr.org/2019/02/crossroads-counter-terrorism-and-the-internet/. For a governmental perspective, see 'How social media is used to encourage travel to Syria and Iraq: Briefing note for schools', UK Depart for Education, July 2015. Accessed at: https://www.gov.uk/government/publications/the-use-of-social-media-for-online-radicalisation. For an international governmental perspective, see 'The use of the Internet for terrorist purposes', United Nations Office on Drugs and Crime, September 2012. Accessed at: https://www.unodc.org/documents/frontpage/Use_of_Internet_for_Terrorist_Purposes.pdf.

password-protected chat forums, such as Ansar al-Mujahideen ('supporters of the mujahideen'), Faloja (a reference to the Iraqi city of Fallujah, which became a hotbed of insurgent activity) and Shamukh ('lofty' or 'someone to be looked up to'), became the primary forms of dissemination for violent extremist content, including videos and communiqués from groups like al-Qaeda, al-Shabaab and Boko Haram.[15]

These forums were relatively static and insular environments. Thus violent extremist content had to be deliberately sought out as it existed in somewhat harder to reach corners of the internet. By the time of the Arab uprisings in 2011, social media had become the dominant form by which violent actors sought both to disseminate content and to win new recruits. The most dramatic encapsulation of this came with the rise of IS and the al-Qaeda aligned Jabhat al-Nusra between 2011 and 2016.[16] The problem was not just their presence on these platforms but the fact that extremist content was now so easily available to anyone who wanted it – with a number of people encountering it purely accidentally. Consequently, the issue of how to address this became particularly acute for tech companies, law enforcement and counter-terrorism policymakers.[17]

Our paper aims to address this by exploring the ways in which automation might help with the identification and disaggregation of this content, allowing tech companies to identify such content more easily when it sits alongside material posted by legitimate users of their platforms.

Before proceeding, it is worth noting that the tools we have developed here have potential implications for other, non-jihadist types of extremist content too, which individual companies will want to monitor based on their own needs. Harmful content online can take many forms, including abuse, bullying, harassment, threats, misogyny, hate speech and terrorist or violent propaganda, among others. These online activities can manifest themselves in real-world harms, although perhaps none so dramatically as that relating to jihadist violence.[18] In any case, they too require the kind of nuanced, contextualised moderation of which we speak below.

15 Evan Kohlmann, 'A beacon for extremists', CTC Sentinel, February 2010, vol. 3, issue 2. Accessed at: https://ctc.usma.edu/a-beacon-for-extremists-the-ansar-al-mujahideen-web-forum/; Manuel R. Torres-Soriano 'The Hidden Face of Jihadist Internet Forum Management: The Case of Ansar Al Mujahideen', Terrorism and Political Violence vol. 28, issue 4 (2016).

16 Gunnar J. Weimann, 'Competition and Innovation in a Hostile Environment: How Jabhat Al-Nusra and Islamic State Moved to Twitter in 2013–2014', Studies in Conflict & Terrorism, vol. 42 (2019): 1–2, 25-42.

17 On the tech side of the equation, see: Fishman, 'Crossroads'. For a governmental perspective, see 'How social media is used'. For an international governmental perspective, see 'The use of the Internet for terrorist purposes'.

18 In late 2018, just four months before its last vestiges in Syria were liberated by coalition-backed forces, IS's use of social media was considered to pose a direct 'threat to stability in the Middle East and Africa'. Antonia Ward, 'ISIS's use of social media still poses a threat to stability in the Middle East and Africa', RAND Corporation, 11 December 2018. Accessed at: https://www.rand.org/blog/2018/12/isiss-use-of-social-media-still-poses-a-threat-to-stability.html.

2 Literature Review

There is a wide body of literature that already looks at violent extremist content online, especially that of IS. After its emergence in Syria and Iraq, IS quickly established itself as a pioneer in both extremist strategic communications and internet-based outreach. Scholarship analysing its output in this regard has usually manifested itself in three different ways: (i) quantitative analysis of its social media support-base; (ii) qualitative analysis of individual propaganda texts or genres; and (iii) data-based analysis of aggregate media output.

The first cluster explores how IS supporters interact with one another online. Since 2014 in particular, their activism on mainstream platforms, such as Twitter, Facebook and YouTube, has attracted a lot of attention. Carter, Maher and Neumann's investigation was one of the first efforts to map these diverse communities, exploring online networks of influence among English-speaking jihadists. This was followed by similarly orientated explorations by respected colleagues, such as Klausen, Berger and Morgan.[19] Subsequent research on the same issue by Conway and others, as well as Alexander, illustrates that the jihadist presence on mainstream platforms declined from 2015 onwards, with newer, privacy-maximising services taking their place as preferred communication hubs.[20] Despite this migration, Winterbotham and others have also shown that mainstream platforms, such as Twitter and Facebook, continue to hold ongoing importance to the movement, even as its activities there declined.[21] Additionally, there are a number of studies that track idiomatic dynamics within extremist communities on social media, using linguistic modelling to detect the presence of radicalised discourse and, on occasion, forecast behaviours.[22] These studies are largely experimental at the moment and do not focus on specific groups.

19 Joseph A. Carter, Shiraz Maher and Peter R. Neumann, '#Greenbirds: Measuring Importance and Influence in Syrian Foreign Fighter Networks', International Centre for the Study of Radicalisation, April 2014. Accessed at: https://icsr.info/wp-content/uploads/2014/04/ICSR-Report-Greenbirds-Measuring-Importance-and-Infleunce-in-Syrian-Foreign-Fighter-Networks.pdf; Jytte Klausen, 'Tweeting the jihad: Social Media Networks of Western Foreign Fighters in Syria and Iraq', Studies in Conflict & Terrorism, vol. 38:1: 1–22; J. M. Berger and Jonathon Morgan, 'The ISIS Twitter census: Defining and describing the population of ISIS supporters on Twitter', The Brookings Project on U.S. Relations with the Islamic World, no. 20, March 2015. Accessed at: https://www.brookings.edu/wp-content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf.

20 Maura Conway et al., 'Disrupting Daesh: Measuring Takedown of Online Terrorist Material and Its Impacts', Studies in Conflict & Terrorism, vol. 42, 2019. Accessed at: http://doras.dcu.ie/21961/1/Disrupting_DAESH_FINAL_WEB_VERSION.pdf; Audrey Alexander, 'Digital Decay: Tracing Change Over Time Among English-Language Islamic State Sympathizers on Twitter', Program on Extremism, George Washington University, October 2017. Accessed at: https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/DigitalDecayFinal_0.pdf; Miron Lakomy, 'Mapping the online presence and activities of the Islamic State's unofficial propaganda cell: Ahlut-Tawhid Publications', Security Journal, online only.

21 Ugur Kursuncu et al., 'Modeling Islamist Extremist Communications on Social Media Using Contextual Dimensions: Religion, Ideology and Hate', Proceedings of the ACM on Human-Computer Interaction, 3:1, August 2019; Moustafa Ayad, ''The Baghdadi Net': How a Network of ISIL-Supporting Accounts Spread across Twitter', Institute for Strategic Dialogue, November 2019. Accessed at: https://www.voxpol.eu/download/report/E28098The-Baghdadi-NetE28099-How-A-Network-of-ISIL-Supporting-Accounts-Spread-Across-Twitter.pdf; Leevia Dillon et al., 'A comparison of ISIS foreign fighters and supporters social media posts: an exploratory mixed-method content analysis', Behavioural Sciences of Terrorism and Political Aggression, online only; Airbus Defence and Space, 'Mapping Extremist Communities: A Social Network Analysis Approach', NATO Strategic Communications Centre of Excellence, January 2020. Accessed at: https://www.voxpol.eu/download/report/web_stratcom_coe_mapping_extremist_strategies_31.03.2020_v2.pdf.

22 Tom De Smedt et al., 'Automatic Detection of Online Jihadist Hate Speech', Computation and Language vol. 7:1–31, 2018; Adam Bermingham et al., 'Combining Social Network Analysis and Sentiment Analysis to Explore the Potential for Online Radicalisation', 2009 International Conference on Advances in Social Network Analysis and Mining, 2009: 231–6; Edna Reid et al., 'Collecting and Analyzing the Presence of Terrorists on the Web: A Case Study of Jihad Websites', International Conference on Intelligence and Security Informatics, 2005: 402–11; Enghin Omer, 'Using machine learning to identify jihadist messages on Twitter', Uppsala Universitet, 2015. Accessed at: http://www.diva-portal.org/smash/get/diva2:846343/FULLTEXT01.pdf.

The second body of work consists of qualitative interrogations of individual propaganda products and genres. There have been myriad explorations into IS's foreign language magazines in recent years, with some also focusing on its official Arabic language newspaper, *al-Naba*.[23] Scholars, such as Winkler and others, as well as Adelman, have preferred to focus on the hundreds of infographics IS published since 2015, while others such as Nanninga, Dauber and Robinson have instead concentrated on its video production.[24] El Damanhoury and Milton are among the few to have examined the expansive archive of IS still images, about which much more can and should be said.[25] Notwithstanding the diversity of their subject matter, these genre-based studies tend to reach similar conclusions regarding the dominant presence of Western visual motifs in IS propaganda.

The last cluster is characterised by work from scholars such as Zelin, Milton and Winter, whose respective efforts revolve around archival studies of official IS media output.[26] Generally speaking, their findings are consistent with one another; they each identify a net decline in the amount of propaganda produced by the group, correlating roughly with its territorial contractions since 2015, something also noted by Nanninga. It must be stressed, however, that this decline was not necessarily caused by the loss of territory, even though the two seem to correlate. While a series of intuitive conclusions may be reached about this, there remains no definitive agreement as to what exactly caused the deceleration and, unsurprisingly, IS has never addressed the matter.[27]

23 Haroro J. Ingram, 'An analysis of Islamic State's Dabiq Magazine', Australian Journal of Political Science, vol. 51:3, 2016: 458–577; Julian Droogan and Shane Peattie, 'Mapping the thematic landscape of Dabiq magazine', Australian Journal of Political Science, vol. 71:6, 2017: 591–620; Haroro J. Ingram, 'An analysis of Inspire and Dabiq: Lessons from AQAP and Islamic State's propaganda war', Studies in Conflict & Terrorism, vol. 40:5, 2017: 357–75; Nuria Lorenzo-Dus et al., 'Representing the West and 'non-believers' in the online jihadist magazines Dabiq and Inspire', Critical Studies on Terrorism, vol. 11:3, 2018; Carol K. Winkler et al., 'The medium is terrorism: Transformation of the about to die trope in Dabiq', Terrorism and Political Violence, vol. 31:2, 2019; Peter Wignell et al., 'Under the shade of AK47s: a multimodal approach to violent extremist recruitment strategies for foreign fighters', Critical Studies on Terrorism, vol. 10:3, 2017: 429–52; Logan Macnair and Richard Frank, 'Changes and stabilities in the language of Islamic state magazines: Aa sentiment analysis', Dynamics of Asymmetric Conflict, vol. 11:2, 2018: 109–20; Orla Lehane et al., 'Brides, black widows and baby-makers; or not: an analysis of the portrayal of women in English-language jihadi magazine image content', Critical Studies on Terrorism, vol. 11:3, 2018; Dounia Mahlouly and Charlie Winter, 'A Tale of Two Caliphates: Comparing the Islamic State's Internal and External Messaging Priorities', VOX-Pol, 2018. Accessed at: https://www.voxpol.eu/download/vox-pol_publication/A-Tale-of-Two-Caliphates-Mahlouly-and-Winter.pdf; Miron Lakomy, 'Towards the 'olive trees of Rome': Exploitation of propaganda devices in the Islamic State's flagship magazine 'Rumiyah'', Small Wars & Insurgencies vol.31:3, 2020: 540–68; Michael Zekulin, 'From Inspire to Rumiyah: does instructional content in online jihadist magazines lead to attacks?', Behavioural Sciences of Terrorism and Political Aggression, online only.

24 Pieter Nanninga, 'Meanings of savagery', in Lewis, J. (ed.), The Cambridge Companion to Religion and Terrorism, Cambridge: Cambridge University Press, 2017: 172–90; Cori E. Dauber and Mark Robinson, 'ISIS and the Hollywood visual style', Jihadology, 6 July 2015. Accessed at: https://jihadology.net/2015/07/06/guest-post-isis-and-the-hollywood-visual-style/; Cori E. Dauber et al., 'Call of Duty: Jihad – How the Video Game Motif has Migrated Downstream from Islamic State Propaganda Videos', Perspectives on Terrorism vol. 13:3, 2019; Pieter Nanninga, 'Branding a Caliphate in Decline: The Islamic State's Video Output (2015–2018)', International Centre for Counter-terrorism – The Hague, April 2019. Accessed at: https://icct.nl/publication/branding-a-caliphate-in-decline-the-islamic-states-video-output-2015-2018/.

25 Kareem El Damanhoury et al., 'Examining the military-media nexus in ISIS's provincial photography campaign', Dynamics of Asymmetric Conflict, vol. 11:2, 2018: 89–108; Daniel Milton, 'Fatal attraction: Explaining variation in the attractiveness of Islamic State propaganda', Conflict Management and Peace Science vol. 37:4, 2018; Carol Winkler et al., 'Intersections of ISIS media leader loss and media campaign strategy: A visual framing analysis', Media, War & Conflict, 2019, online only.

26 Aaron Y. Zelin, 'Picture Or It Didn't Happen: A Snapshot of the Islamic State's Media Output', Perspectives on Terrorism vol. 9:4, 2015; Daniel Milton, 'Communication Breakdown: Unraveling the Islamic State's Media Efforts', Combating Terrorism Center at West Point, 2016. Accessed at: https://ctc.usma.edu/communication-breakdown-unraveling-the-islamic-states-media-efforts/; Daniel Milton, 'Down, but Not Out: An Updated Assessment of the Islamic State's Visual Propaganda', Combating Terrorism Center at West Point, 2018. Accessed at: https://ctc.usma.edu/down-but-not-out-an-updated-examination-of-the-islamic-states-visual-propaganda/; see also: Charlie Winter, 'Apocalypse, later: A longitudinal study of the Islamic State brand', Critical Studies in Media Communication, vol. 35:1, 2018: 103–21.

27 It is worth noting that the consensus is not quite complete, with an outlier account by Fisher contending that there has been no such productivity decline. Ali Fisher, 'ISIS: Sunset on the "decline narrative"', Online Jihad, 2018. Accessed at: https://onlinejihad.net/2018/06/01/isis-sunset-on-the-decline-narrative/.

The present study builds on the first and third bodies of work. In relation to the first, it offers a new form of experimental text processing, focusing not on social media posts but on the content itself. Its contribution to the third cluster is easily apparent, given the archival and multimedia nature of the data. Through it, we hope to lend further nuance to the debate around how and why IS's outreach activities have evolved in recent years.

3 Methodology

The most important aspect of our methodology relates to the way in which the automated text analysis tool was developed and deployed. This is the premise on which we have predicated our model of automatically parsing content at scale in an attempt to triage potentially harmful material.

The dataset we used was drawn from a static website administered by an unknown IS supporter (or supporters).[28] Easily accessible on the surface web, the site has been in circulation among jihadists for a number of years, with links appearing on public discussion boards, mainstream social media platforms and more inward-looking platforms such as Telegram. The website was downloaded, archived and stored in its entirety in February 2020, totalling 6,290 individual items – anything from photo-reports and videos to leadership statements, radio bulletins, and magazines.

Developing our Thematic Framework

The algorithm we developed (which is explained below) was based on an analytical framework designed to disaggregate IS content based on different themes, developed by one of the authors, Dr. Winter, for a previous project.[29] This framework contains two categories – (i) war and (ii) civilian life – under which 22 different thematic categorisations are found.[30] These are listed below, with nine relating to war and thirteen relating to civilian life.

(i) War themes

1. *Operations*: Offensive military operations. These vary according to context, stated objective and tactic(s) employed. The three most frequently appearing iterations depict ground assaults, improvised explosive device (IED) operations and suicide attacks. Other tactics that feature include anything from drone strikes to night-time ambushes.

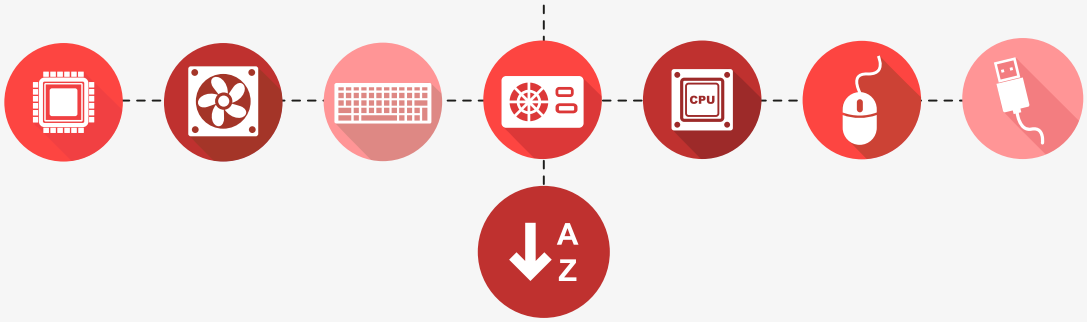
2. *Summary*: Aggregated news reports from across the caliphate territories. They manifest in daily, weekly and monthly digests, often supported by statistics.

3. *Indirect warfare*: Attacks against enemy positions using rockets, missiles and mortars. These tend to focus on the launching of projectiles and rarely show the aftermath of the strikes themselves.

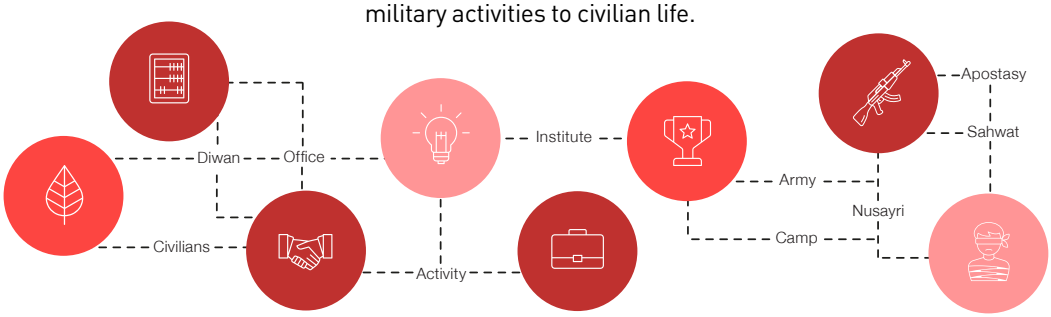
28 We have opted not to name the archive in question because it remains easily accessible online and doing so would needlessly advertise it. Readers are requested to contact the authors directly with any queries.
29 Charlie Winter, 'The Terrorist Image: A Mixed Methods Analysis of Islamic State Photo-Propaganda', Unpublished PhD thesis, King's College London, July 2020.
30 Also see: Milton, 'Communication Breakdown'; Milton, 'Down, but Not Out'; Zelin, 'Picture Or It Didn't Happen'; Winter, 'Apocalypse, later'.

The archive was downloaded and secured.

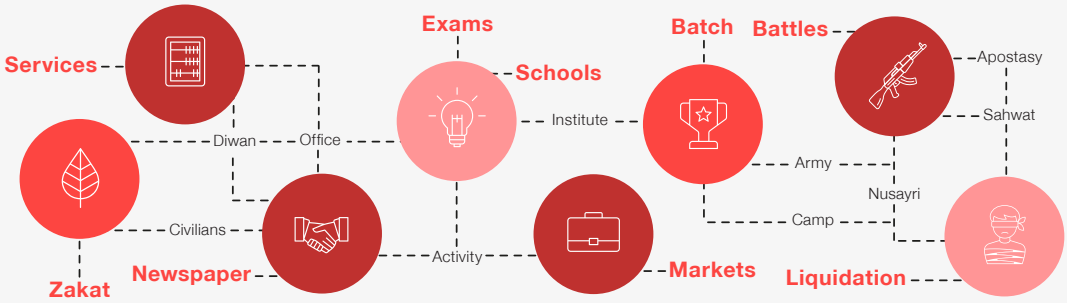
The 6,290 items it contained were sorted by media type and date of publication.
The 803 words that appear most frequently in the titles of each were identified.



These words were then tagged with one to three themes, covering anything from military activities to civilian life.



When words related exclusively to one theme, they were designated super-tags.



The algorithm applied this system of tags and super-tags to the whole dataset.

4. *Martyrology*: These materials glorify men and boys that die fighting in the name of IS. Almost all of the individuals eulogised in these images are photographed while they are alive, either in idyllic bucolic or urban surroundings or inside the vehicle-borne bombs in which they are set to die. Those they commemorate range from suicide operatives to low- or mid-level leaders and propaganda officials.

5. *Garrison duty*: Materials aggrandising life on the IS front lines. In the main, they give an overview of quotidian goings-on in the garrisons, focusing on anything from prayer time and food preparation to weapons cleaning and physical exertion.

6. *Executions*: Materials documenting the execution of 'spies' or prisoners of war captured during kidnappings or assaults. This includes members of rival violent extremist groups. Propaganda relating to executions occurring in an overtly martial context should not be confused with propaganda relating to executions occurring in a civilian context.

7. *Defensive operations*: Defensive military activities. Most of these materials revolve around foiled enemy offensives and 'successful' counterattacks. Others relay information about anti-aircraft defences, military preparedness measures, the building of fortifications and maintenance of weapons systems.

8. *Aftermath*: The aftermath of an attack. In this category, there are four clear groupings: war spoils; enemy captives; enemy corpses; and damaged ground vehicles and drones.

9. *Training*: These materials depict training camps, showcasing activities like weapons drills, physical exercise and hand-to-hand combat.

(ii) Civilian life themes

10. *Law and order*: Administration of law and order in IS territories. These materials manifest in three main ways: images of religious policing (*hisbah*); images of penal proceedings; and images of police deployments.

11. *Victimisation*: These materials document the aftermath of attacks against IS territories, typically featuring dead or injured children and devastated public infrastructure. They are used to justify IS's rule, win support for the group and instigate violent action in response.

12. *Outreach*: Most of these materials follow the work of media officials, principally their dissemination of content and setting up of media infrastructure. Others showcase activities like reconciliation meetings between rival tribes and gatherings with civilians and dignitaries.

13. *Tours*: These materials document 'everyday' life inside IS strongholds. Usually depicting structured visits to specific villages, towns or neighbourhoods, content in this category tends to track anything from religious activities and birdlife to trade and recreation.

14. *Religious life*: Materials mostly focusing on religious activities, these show civilians in a range of 'Islamic' contexts, with anything from Eid and Ramadan festivities to Friday prayers and Quran memorisation contests.

15. *Commercial life*: Materials showcasing commercial aspects of life inside the caliphate. Most focus on trade and commerce, with others depicting tours of markets, shops and showrooms.

16. *Municipal services*: The provision of services in IS territories. Highly varied, they show service bureaus engaging in anything from pylon repairs to sewer maintenance.

17. *Social welfare*: These materials depict the calculation, preparation and distribution of social welfare (financial and food-based) among the civilian population of IS-held territories. Most offer 'day in the life'-style vignettes giving a general overview of the activities of the welfare office.

18. *Industrial life*: Materials focusing on industrial activity in the caliphate, including anything from pipe factories and air-conditioner workshops to cheese production plants and sunflower-seed drying facilities.

19. *Agricultural life*: These materials cover agricultural activity, including anything from seeds being sown to fruit being harvested and taken to market.

20. *Education*: These typically revolve around seminary activities put on by official proselytisation authorities. Others depict school visits and include anything from teenagers completing mid-term and end-of-year exams to infants playing games during break-time.

21. *Healthcare*: These materials showcase medical activities under IS. They range from tours of hospitals and dental clinics to home visits from general practitioners and vaccination campaigns for children.

22. *Landscapes and nature*: This content mostly consists of photographs of sites of natural or monumental beauty.

Developing the Algorithm

Our algorithm was coded to look for linguistic markers present in the title alone of each of the 6,290 individual items identified in our database. Many millenarian and reactionary movements employ their own lexical and syntactic markers of language, which define the in-group. One of the most prominent examples of such markers can be found in the uses of triple brackets around the (((names))) of prominent Jewish individuals by members of some alt-right and neo-Nazi communities. The original idea behind this is to identify Jewish people or those of a Jewish background.[31] Those associated with the alt-right and/or neo-Nazi movement would then know to view this individual from within the prism of antisemitic conspiracy or intrigue.

Another example of language specific to certain online subcultures can be found within the incel ('involuntary celibate') community, which comprises of sexually inactive men who are hostile to, and distrusting

31 Matthew Yglesias, 'The (((echo))) explained', Vox, 6 June 2016. Accessed at: https://www.vox.com/2016/6/6/11860796/echo-explained-parentheses-twitter.

of, both women and feminism.[32] The term 'Chad', for example, is used by incel communities as a derogatory reference to attractive, popular men who are assumed to be sexually active with women. Similarly, the term 'Stacy' is used to denote attractive, beautiful women who are only interested in 'Chads'.[33]

The same is true for IS's titling of its content. It uses linguistic markers to denote particular themes or issues. Hence, if one is familiar with the requisite language, then it is possible to make judgements about the thematic nature of content based on title alone. This approach is not infallible, of course, but it is sufficient for our goal of finding ways to identify, classify and disaggregate extremist content in a triaged way for subsequent human review. We therefore directed and augmented our text processing tools on this basis.

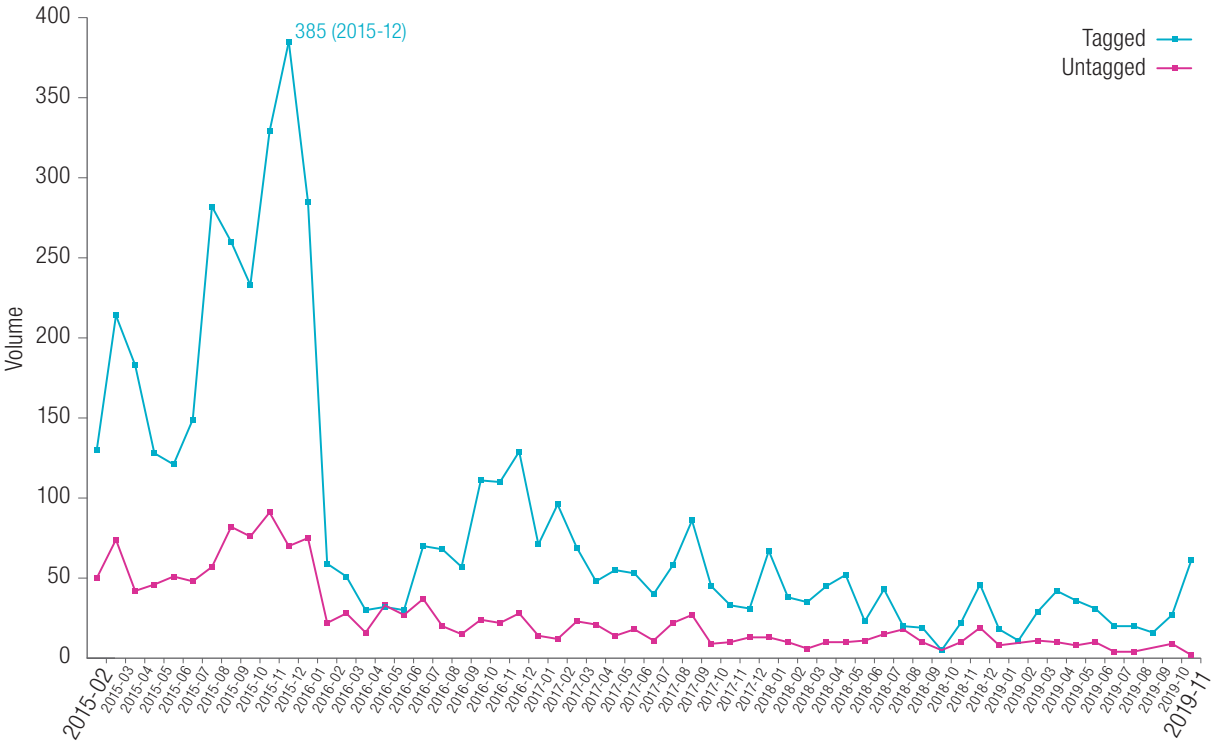
Prior to developing the algorithm, however, we first had to ensure that our framework was valid. To do this we created a codebook and tested it for intercoder reliability.[34] This was done by selecting a random sample of 286 items (that is, 5% of the overall material) and having three researchers from ICSR then independently code the material. All were Arabic speakers and familiar with IS content. In all but 41 cases (14% of the overall sample) the items were coded consistently, with the remainder then being individually discussed and reconciled between the coding cohort. Most of these inconsistencies stemmed from translational disagreements over the precise syntactic or contextual application of a specific Arabic term.

Having calibrated the framework through the intercoder process, we then programmed our algorithm to categorise the entire cache of archived material automatically. The algorithm was taught to do this by associating specific words with specific themes, a process visualised in the infographic on page 12. With regard to our dataset, we first identified the 803 words that recurred with the greatest frequency, in order to achieve full 'coverage' of the corpus – i.e., to ensure that all of its 6,290 items were accounted for by this part of the process. We arrived at this number based on a process of trial and error to find the most efficient way to provide overall coverage of the archived cache. These 803 words, which excluded place names and proper nouns, were then associated with one to three themes from the 22 listed in our framework.

Words that appear in the context of more than one theme were tagged as such. For example, the word *mahud* ('institution') was associated with the themes 'Education', 'Training' and 'Outreach' because it is variously used to describe IS schools, military camps and religious seminaries. Similarly, *riddah* ('apostasy') was tagged with the themes 'Operations', 'Executions' and 'Law and Order', because it is used in the context of all three. By tagging the 803 words in this way, we were able to assign all 6,290 items in the corpus with between one and three themes.

32 Rebecca Jennings, 'Incels categorize women by personal style and attractiveness', Vox, 28 April 2018. Accessed at: https://www.vox.com/2018/4/28/17290256/incel-chad-stacy-becky.
33 'A parent's guide to the secret language of internet extremists', CBS News, 16 March 2020. Accessed at: https://www.cbsnews.com/news/incels-radicalization-glossary-parents-cbsn-orginals-extremists-next-door/.
34 Moin Syed and Sarah Nelson, 'Guidelines for Establishing Reliability When Coding Narrative Data', Emerging Adulthood vol. 3:6, 2015; Paul J. Lavrakas, Encyclopedia of Survey Research Methods, Thousand Oaks, CA: Sage Publications, 2008.

Figure 1: Coded items vs. uncoded items



The knowledge that an item potentially relates to one to three divergent themes is not especially useful, so we then imposed a series of 'super-tags' that were overlaid on top of this initial process. Each super-tag constitutes a single word that is used exclusively in the context of a specific theme. For example, the term *aswaq* ('markets') was assigned a super-tag for the theme 'Commercial life' because it appears solely in the context of commercial activities and nothing else. If, say, the words 'institution' and 'markets' were to appear together in the same title, the presence of the super-tag 'Markets' would mean the item would automatically be classed as 'Commercial-life'-themed, regardless of what other words were present. In total, 232 super-tags were created.

Once instituted on top of the initial text analysis, which was based on the 803 linguistic identifiers, the super-tag system meant that 4,848 of the items (79% of the archived material) were successfully assigned a single theme. We then checked the results produced by the algorithm against those produced by our human researchers. The algorithm was found to be in agreement with our human teams' fully reconciled coding 91% of the time. While still imperfect, we judged this margin of error to be sufficient for present investigation, which is intended to be exploratory in nature.

The algorithm could not code the remainder of the archive, which consisted of 1,432 items. This primarily consisted of audio statements and videos, the titles of which tend not to be descriptive in nature, instead quoting Islamic scripture or other, more obscure sources. This meant they did not contain any of the 803 linguistic identifiers or 232 super-tags and were consequently not assigned a theme.

The algorithm also encountered problems when it handled titles that contained words relating to more than one theme, but no super-tags. For example, a photo-report entitled ‘the production of fishing boats’ contains a marker ‘fishing’, which is linked to the theme ‘Agricultural life’, in addition to another, ‘production’, which is linked to the theme ‘Industrial life’. Because both of these markers were present and there was no super-tag, the algorithm could not apply a single theme to the item in question.

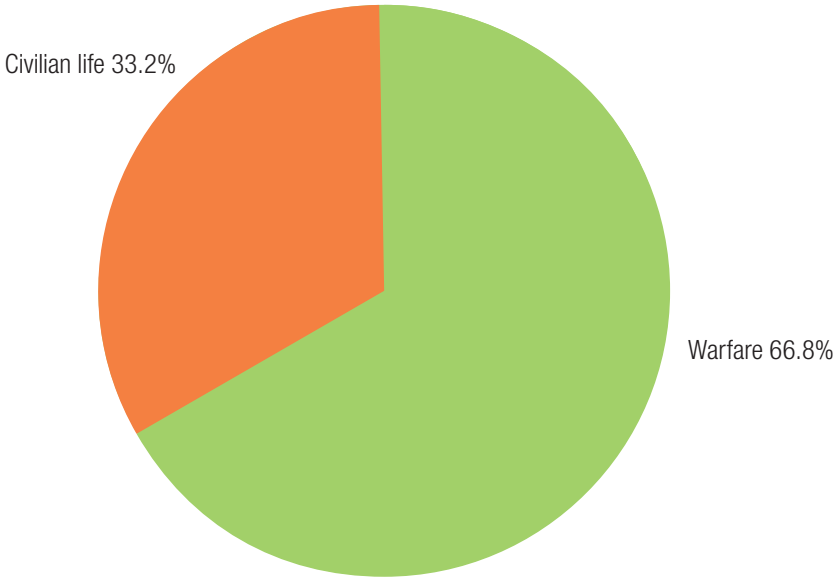
Figure 1 shows items that were successfully coded by the algorithm and items that were not. The correlation of the two lines indicates that the algorithm was equally valid throughout the period in question. Our study shows that nearly a fifth of the corpus went uncategorised, constituting a significant limitation to our approach. Its resolution, which would involve an entirely new set of methodological tools, requires exploration of factors other than linguistic analysis that are beyond the scope of this study. We will turn to this in the future.

4 Findings

Triaging the Data

Once our algorithm concluded its analysis of the 6,290 items in our dataset, we discovered that it was able to help with the identification and triaging of material. As Figure 2 shows, 66.8% of the archive consisted of war-related themes while just 33.2% was about civilian themes. At a practical level this has clear implications for technology companies looking to parse referred content along even the most basic of initial distinctions: war-related material versus civilan-life-focused material.

Figure 2: Overarching themes of tagged content



This becomes more interesting when we explore what types of war-related material featured most frequently. Figure 3 shows that the most frequently appearing material was 'Operations' which comprised 30.3% of our dataset. This was followed by 'Summary', at 10.4%; 'Indirect warfare', at 7.2%; and 'Executions' at 3.2%. Thus, it can be seen within this category that although executions are not the most prominent type of violent material, a company may nonetheless still wish to prioritise this particular type of content for more urgent review given the nearly always graphic and harmful nature of this content. The other categories, although more prevalent and still pertaining to war, are still likely to contain a greater frequency of non-graphic material, such as military hardware, weaponry and related items.

It is certainly not our aim to suggest that technology companies are employing crude or blunt instruments with which to detect violent extremist content on their platforms. Clearly this is not the case and

a great array of complex instruments work to identify harmful content. The purpose of this algorithmic tool is to complement those existing practices by employing a process of linguistic-marker identification to further enhance the triaging process.

Our results, visualised in Figure 4, also show that of the civilian-life-focused propaganda the most prominent themes are: 'law and order',

Figure 3: Themes in war-focused propaganda

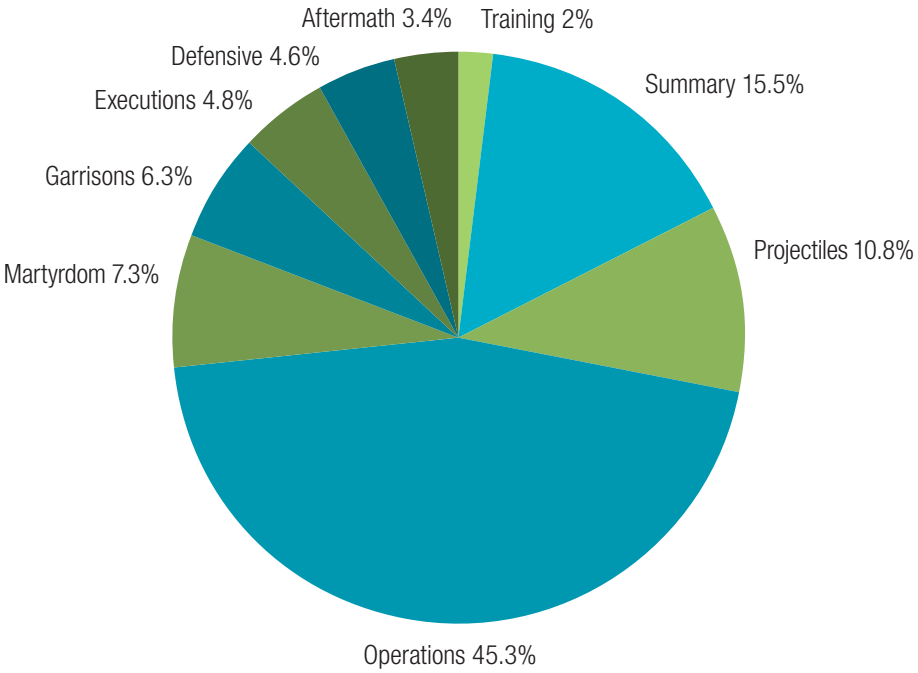
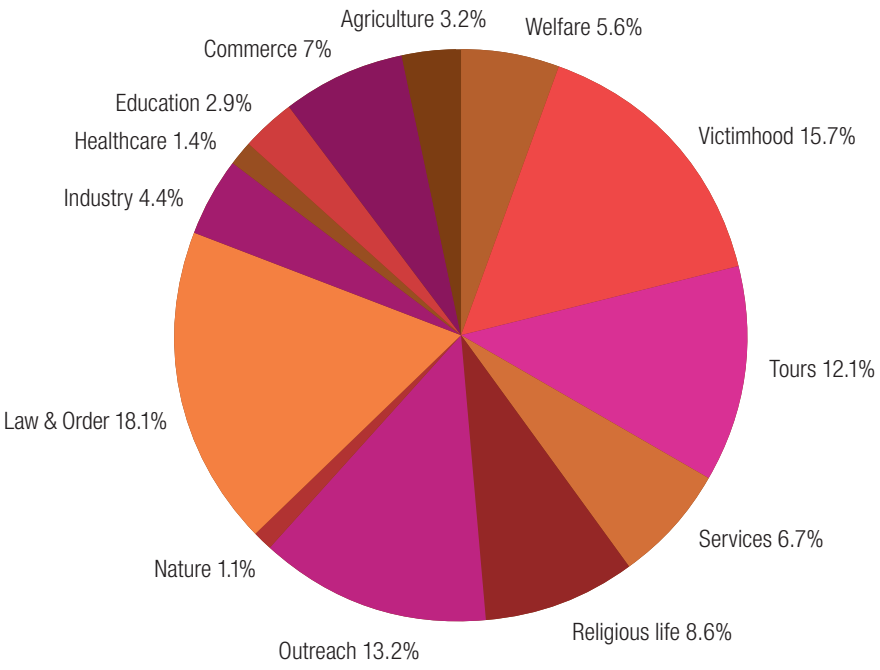


Figure 4: Themes in civilian-life-focused propaganda



at 6%; 'victimhood', at 5.2%; 'outreach', at 4.4%'; and 'tours', at 4%. Given that IS content focusing on law and order routinely features both benign scenes of police patrols and violent scenes of accused criminals being executed and/or maimed, it should follow that it is prioritised for urgent human review immediately after the war-related 'executions' theme. Immediate review is somewhat less of a concern for other civilian-life-focused content because, with the exception of victimhood-related propaganda, which revolves around civilian casualties, these materials are rarely if ever violent.

The algorithm also allows us to detect and visualise changes in thematic emphasis or prioritisation over time as shown in Figure 5. This shows a gradual move away from civilian-life-focused materials, which comprised around 49% of items listed in 2015 but just 7% of those listed in 2019. This reflects changes on the ground, as IS lost territory and shifted its messaging away from showcasing the apparent comforts of its supposed utopia, towards material that is angrier and more war-focused, echoing traditional jihadist tropes of hostile 'Crusader powers' waging a 'war on Islam'. The aim of its propaganda therefore shifted from suggesting supporters should migrate to Syria and Iraq to support their 'caliphate' to encouraging them to stay at home and conduct terrorist attacks there.

During this time, there was a gradual simplification of IS's overall discourse too. Indeed, as its productivity declined through the period 2016 to 2019, so too did the thematic variance of its output. This bears out in the data. As Figures 6 and 7 show, in 2015, its output was a composite of all 22 of the categories described above. By 2019, however, the output comprised just 14 of them – nine of which were war-related.

Figure 5: Thematic priorities of tagged content (by month)

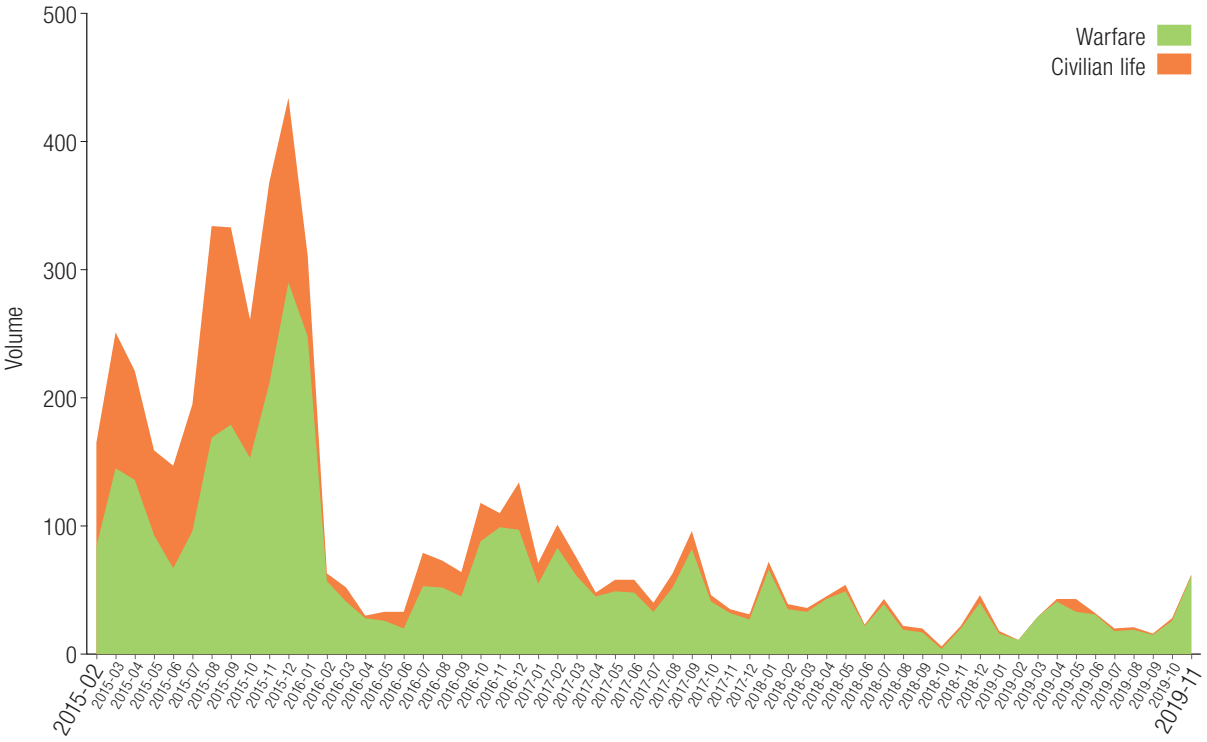


Figure 6: War-focused and civilian-life-focused content in 2015

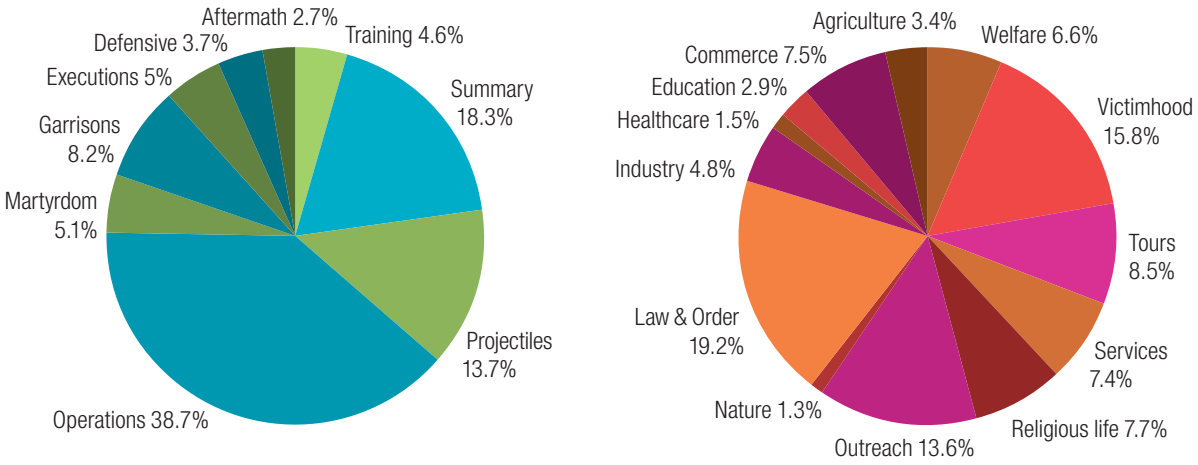
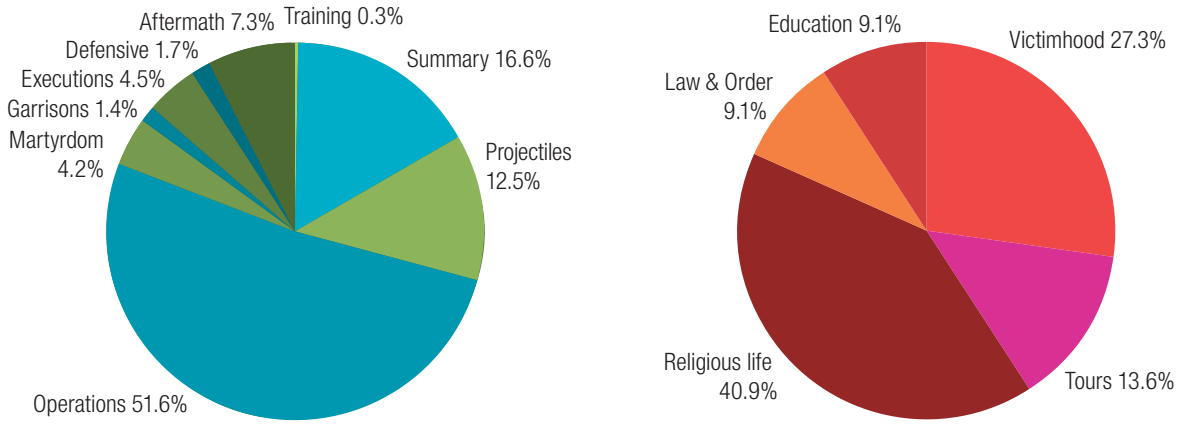


Figure 7: War-focused and civilian-life-focused content in 2019



This simultaneous transformation and simplification of IS's narrative is by no means confined to the archive. Rather, it is a dynamic that characterises how IS's official outreach activities altered to accommodate its new situational exigencies on the ground. Simply put, as its territories declined and the nature of its war changed course, IS entered into a new paradigm of insurgent warfare. Instead of fighting for material and territorial gain, its operational roster reverted to activities that were primarily geared towards progressive, underground consolidation. To this end, the tenor of its propaganda changed, becoming more about signalling resolve and demonstrating sustained presence than accruing new cadres or provoking global outrage. As such, as tracked in Figures 5, 6 and 7, its content focused less on civilian life in IS's proto-state and more on IS's war effort.

This visualisation aspect of our algorithm can therefore allow technology companies to monitor changes in emphasis and priority over time, giving them the ability to adapt if they are able to better anticipate more of one particular type of material than another.

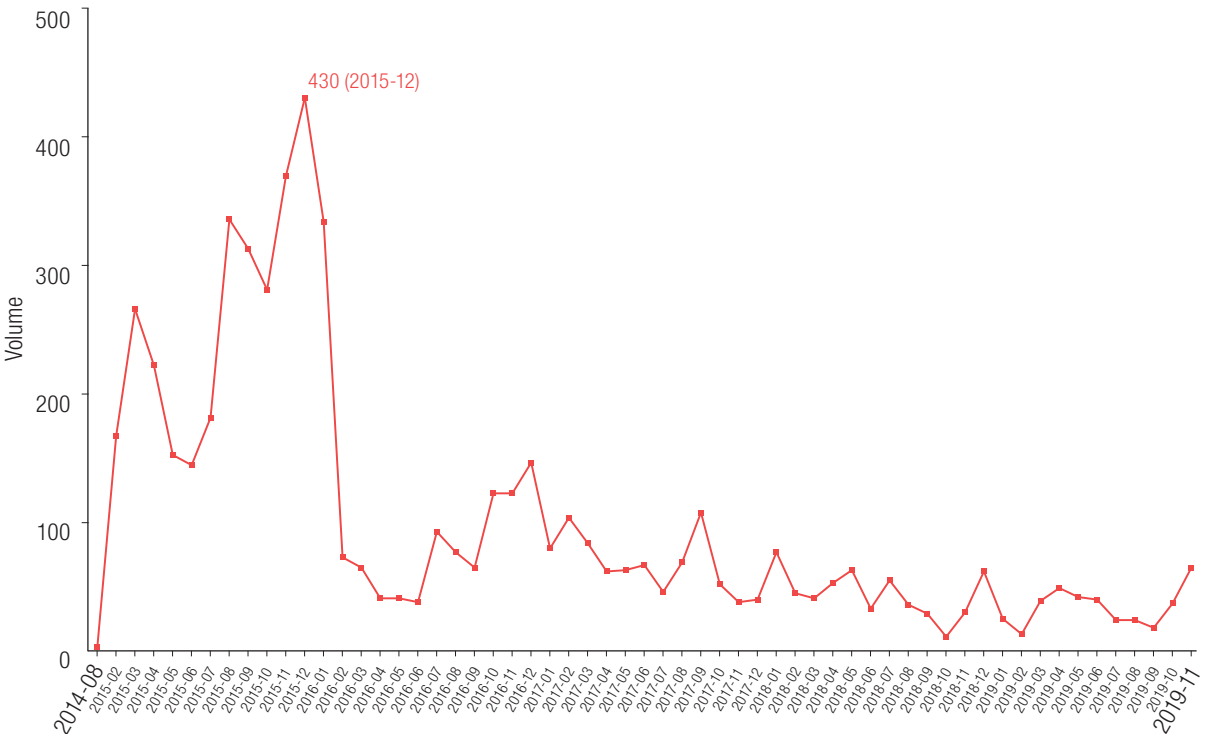
Utility of Identifying Temporal Characteristics

Figure 8 shows the temporal distribution of all items in the archive. Each of the 6,290 items was sorted according to its date of publication and ordered consecutively. This allowed us to better understand and visualise the amount of content that was published over time for this dataset. In a live setting this visualisation would be ongoing, changing in real-time, but is provided here for demonstration purposes.

The graph indicates that most of the content is drawn from 2015, after which there is a precipitous drop that broadly continues through the period 2016–19. This probably tells us more about the partial nature of the archive than anything else. As we know, IS's production of propaganda tailed off significantly in the period 2017–19. Hence, we can be confident that the archive is more complete for these years. Conversely, for 2015–16, even though the archive captures a significant amount of propaganda, that in itself is only a small proportion of what was produced during the period. Moreover, it contains just one item from 2014, a year in which many thousands of pieces of content were published by IS. The most likely explanation for this discrepancy in coverage is that whoever created the archive began creating it after most official IS content had already been removed from the surface web. If this is the case, then it would follow that they were only able to capture a small proportion of the official materials that emerged during the years in which they were interested.

Whatever the case, the year-on-year decline shown in Figure 8 remains broadly consistent with how IS's media production capabilities changed after 2015. As a number of researchers have already

Figure 8: Total content volume (by month)



shown, 2015 represented a high watermark in terms of IS production of propaganda.[35]

This decline confirms that territorial control correlated strongly with content production capabilities for IS. When the group presided over a population of millions and was engaging in various forms of governance, its propagandists were not only able to enjoy greater freedom of movement and access to monetary and human resources but also were inundated with subject matter about which to craft new material.[36] Additionally, when IS was at its territorial apex, it was simultaneously fighting on more than a dozen fronts in mostly conventional ways.[37] This lends itself to more propagandistic coverage than covert operations, with the latter becoming more prominent when IS was forced onto the defensive.

Changes in the internal composition of IS will also have played a role in precipitating its media decline. Naturally, as the contours of its insurgency changed so too did its outreach priorities, leaving it less focused on recruitment and more interested in the retention of its local support base.[38]

The relevance of this for the present purposes is that our algorithm can potentially aid companies with detecting the efficacy of their takedown efforts. Clearly, some content is more difficult to remove or can be hidden more easily. Numerous studies have shown how IS and its supporters are conscious of the need to disguise both themselves and their content when operating on mainstream platforms.[39] Overwhelmingly, these efforts do not succeed. Nonetheless, our tool can aid with the identification of resilient linguistic markers that are able to persist over time, underscoring the need for this temporal dimension of analysis.

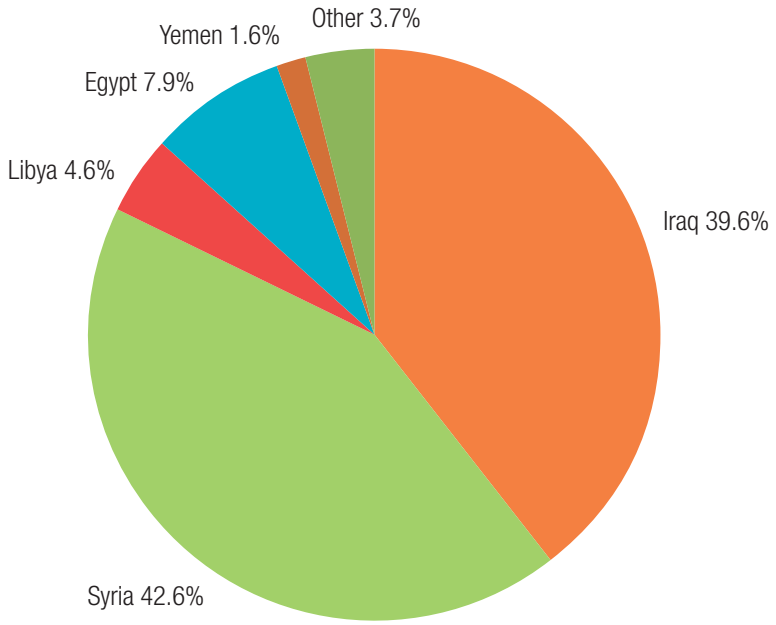
Geographic Characteristics

Given the ability of algorithmic tools to detect insignias, logos and other forms of branding, we also looked for geographic characteristics within our dataset. Again, if done in real-time this would allow for technology companies to train their existing tools to scout for more of one particular type of content than another. This would be especially useful in the context of a foreign fighter mobilisation effort like that deployed by IS between 2013 and 2015 in particular: if content hailing from the destination of said mobilisation (in the abovementioned case, Syria) could be prioritised for removal, tech moderators could meaningfully undermine efforts to advertise the alleged benefits of enlisting.

To do this, we catalogued materials according to the place from which they ostensibly originated. We did this based on the IS media unit that was responsible for producing the content. In the years covered by our archive, IS operated a three-tiered system for media production comprising central offices such as the al-Furqan Foundation and AlHayat Media Center, auxiliary agencies like the Amaq News Agency

35 See, for example: Milton, 'Communication Breakdown'; Milton, 'Down, but Not Out'; Winter, 'Apocalypse, later'; Nanninga, 'Branding a Caliphate in Decline'.
36 Aaron Y. Zelin, 'The Islamic State's Territorial Methodology', The Washington Institute for Near East Policy, 2016. Accessed at: https://www.washingtoninstitute.org/policy-analysis/view/the-islamic-states-territorial-methodology.
37 For an account of its evolution, see Ahmed S. Hashim, 'The Islamic State's Way of War in Iraq and Syria: From Its Origins to the Post Caliphate Era', Perspectives on Terrorism, vol. 13:1, 2019: 23–32.
38 One indication of this is the fact that IS has not published any new video or magazine content through the AlHayat Media Center, its most outward-looking, foreign-language propaganda foundation, since January 2019.
39 See Berger and Jonathon Morgan, 'The ISIS Twitter census'.

Figure 9: Content by state



and Furat Media Center, and provincial media units such as the Wilayat al-Sham and the Wilayat al-Iraq Media Offices.[40] Before the summer of 2018, IS's media network in both Syria and Iraq was subdivided into 23 regional media offices, one for Wilayat al-Raqqah, another for Wilayat Halab, and so on.[41]

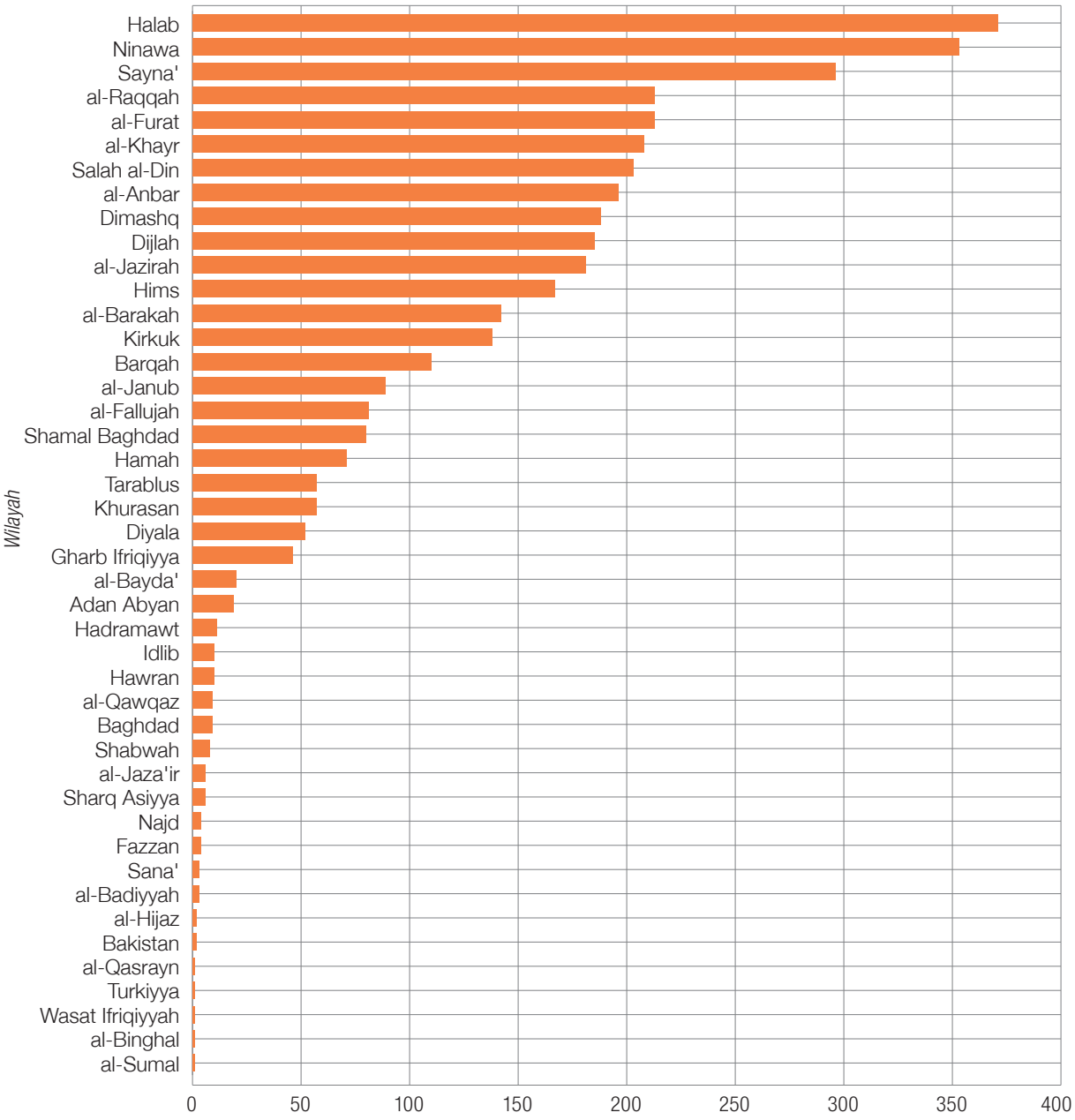
In our archive, 3,831 items could be classed as produced by provincial media offices. A further 554 were identified as being prepared by a central media unit, mainly the al-Furqan Foundation, AlHayat Media Center, al-Itisam Foundation, al-Bayan Radio or al-Naba. The remaining 1,905 items, most of which were published by the Amaq News Agency, could not be classified geographically as no location-specific media office was identified in the archive index.

Figure 9 indicates that, of the items that were tagged with a location-specific media office, 42.6% hailed from Syria. A further 39.6% were produced by IS's media offices in Iraq. Moreover, 7.9% were prepared by the media office for Wilayat Sayna' in Egypt, with the remaining items originating, in descending order, in Libya, Yemen, Afghanistan, Nigeria, Russia, Algeria, Indonesia, Pakistan, Saudi Arabia, Turkey, Tunisia, Democratic Republic of the Congo, Bangladesh and Somalia.

Figure 10 illustrates the items by *wilayah*, not state. It shows that most of the content in the archive was produced by just three of IS's media units: the Wilayat Halab Media Office in Syria; the Wilayat Naynawa Media Office in Iraq; and the Wilayat Sayna' Media Office in Egypt.

40 Abu Abdullah al-Masri, 'The Isis papers: A masterplan for consolidating power', The Guardian, 7 December 2015. Accessed at: https://www.theguardian.com/world/2015/dec/07/islamic-state-document-masterplan-for-power.
41 BBC Monitoring, 'Analysis: The Islamic State restructures its 'provinces' a year on from 2017 defeats', 17 October 2018. Accessed at: https://monitoring.bbc.co.uk/product/c200bdcn.

Figure 10: Content by *wilayah* media office



The prevalence of content from Syria and Iraq is consistent with the broader geographic character of IS during the years in question.[42] However, materials produced by the Wilayat Sayna' Media Office are over-represented in the archive, suggesting that its creator was either more interested in or better able to access materials relating to IS's activities in Egypt. This suggests more about the creator's origins than the provenance of the content itself.

42 See Winter, 'Apocalypse, later'.

5 Conclusion

This research paper used automated text processing techniques to develop a series of analytical tools capable of both analysing and classifying IS propaganda at scale. Although we have looked exclusively at IS material, the ideas established here are, in principle, applicable to any form of violent extremism. Using an archive of 6,290 items as a sample, we attempted to automate the temporal, geographic and thematic categorisation of violent extremist content. Our tool successfully disaggregated the dataset, allowing for identification of the most potentially damaging content over less urgent, although nonetheless highly problematic, content. The broader application of our tool also allowed for the identification of a number of other widely corroborated dynamics, whether in relation to output decline, geographic characteristics or narrative simplification.

Our principal objective was to develop methods that would, when applied to similar corpuses of materials (including those that are much larger), accelerate the process by which they may be disaggregated and, where relevant, triaged for moderation and/or referral.

From the outset, we aimed to produce a tool that could disaggregate content in order to aid the triaging of material for human review. This tool is not, of course, intended to act in a solitary context. We recognise that technology companies have a number of sophisticated systems already in place for the detection and removal of content. In this vast majority of cases this happens in a fully automated context. Yet, human reviewers will also continue to play an important role when adjudicating on more contentious material and this is where our tool is capable of streamlining existing processes.

We believe tools such as this will only become more important given the diversification of challenges facing technology companies. These include state-backed disinformation campaigns, coordinated inauthentic behaviour, and the proliferation of conspiracy theories. Due to high levels of redundancy in the way violent extremist organisations present their wares – they use the same finite set of logos, adopt the same introduction sequences and overlay the same audio-tracks – these materials present relatively low-hanging fruit when it comes to automated detection. However, when it comes to limiting offline harms while upholding the principle of free speech, the ability to detect alone is insufficient. These materials must also be processed and understood both in their broader context and at the level of the intent behind their production. As things stand, this is something that only human moderators are capable of doing.

The tools developed in the course of this investigation would go some way towards streamlining that process of triaging for human review.

Policy Landscape

This section is authored by Armida van Rij and Vivienne Moxham-Hall, both Research Associates at the Policy Institute based at King's College London. It provides an overview of the relevant policy landscape for this report.

Introduction

The misuse of the internet by terrorists and people holding extremist views has become a growing problem over the past decade. Social media platforms and other tech companies' platforms are being used to share terrorist propaganda, recruit people to terrorist organisations or incite violence. Classifying such harmful content or, more simply, deciding what is illegal and should be taken off the platform has proven to be a real challenge for tech companies and policymakers alike. The process inevitably leads to making decisions about what counts as 'extremist' or 'terrorist', about the resources and capacity required to moderate hours of new content uploaded on an hourly basis, and about how one might block harmful content while ensuring freedom of speech, thought and debate. This combination of factors has made tackling online terrorist content and classifying what is and isn't permissible a very difficult challenge, but one with very real consequences. The live-streaming on Facebook of a terrorist attack on two mosques in Christchurch, New Zealand, was viewed at least 4,000 times before being taken down, but this did not prevent the re-uploading and circulating of the footage on social media sites, including Facebook.[43]

To better understand how countries are tackling the challenge of online terrorist content, how to best categorise it and what actions states have taken to remove harmful content, we have assessed the policy landscape of nine legislatures with regard to online terrorist content. They have not been picked at random but are instead members of the Independent Advisory Committee (IAC) for the Global Internet Forum to Counter Terrorism (GIFCT). The IAC itself consists of 21 members, including representatives from seven governments, two international organisations, and 12 civil society organisations (CSO), representing a range of expertise. The relevant legislatures are:

- Government of Canada
- Government of France
- Government of Ghana
- Government of Japan
- Government of New Zealand
- Government of the United Kingdom
- Government of the United States
- European Union
- United Nations Security Council Counter-Terrorism Committee Executive Directorate (CTED)

43 'Facebook: New Zealand attack video viewed 4,000 times', BBC News, 19 March 2019. Accessed at: https://www.bbc.co.uk/news/business-47620519.

For each of these legislatures, this report will address the following questions: (i) Who are the key stakeholders in each legislature?; (ii) What challenges do they face?; (iii) What are the policy developments and key legislation in each jurisdiction; and (iv) What are the key stakeholders planning for the future?

With regards to the first research question, there clearly are myriad stakeholders in each legislature, from government departments and internet service providers (ISPs) to social media companies, civil society organisations and the public. In this report, we will not endeavour to map all stakeholders related to a particular state or organisation. Instead, we will focus solely on the key stakeholders with crucial responsibilities for tackling online extremist content.

Before giving a country-by-country assessment, it is possible to describe a number of challenges shared by all. One challenge countries in the West share in particular is the need to balance the right to freedom of speech with the need to protect populations. Free speech advocates have in the past criticised governments for legislating the removal of extremist videos online, warning that this could infringe on freedom of speech and ultimately amount to censorship.

Additionally, smaller platforms are increasingly being used to host extremist content, but do not have the capabilities to monitor, review and take down illegal content. While the bigger social media and IT companies have more resources and are as a result better equipped to deal with such challenges, this has proven difficult in the past for smaller organisations.

Canada

The Department of Public Safety and Emergency Preparedness (Public Safety Canada) has supported the development of the Terrorist Content Analytics Platform. Statistics Canada, Canada's national statistical office, tracks terrorism in Canada. The Canada Centre for Community Engagement and Prevention of Violence leads Canada's efforts to counter radicalisation, working across government, civil society, law enforcement agencies and international organisations.

In 2017, the Canada Centre launched the National Strategy on Countering Radicalisation to Violence, which focuses on early prevention, at-risk group prevention and disengagement from violent ideologies.[44] Canada currently already uses categorisation to track terrorism. Specifically, Statistics Canada tracks terrorism based on 13 Uniform Crime Reporting (UCR) violation codes.[45] These codes currently do not specify online activity as a specific category. Yet, from a survey of 13 municipal Canadian police agencies conducted in 2016, 40% were unaware of the UCR codes, and about half found it difficult to determine which code to use in particular instances.[46] This highlights the broader challenge states and law enforcement agencies face in defining terrorist activity.

44 See 'Canada: Extremism & counter-extremism', Counter-Extremism Project, 23 June 2020. Accessed at: https://www.counterextremism.com/countries/canada.

45 Patrick McCaffery et al., 'Classification and Collection of Terrorism Incident Data in Canada', Perspectives on Terrorism, vol. 10:5, 2016: 43. Accessed at: https://www.universiteitleiden.nl/binaries/content/assets/customsites/perspectives-on-terrorism/2016/issue-5/505-classification-and-collection-of-terrorism-incident-data-in-canada-by-patrick-mccaffery-lindsy-richardson-jocelyn-j.-belanger.pdf.

46 Ibid.

Justin Trudeau, Canada's prime minister, joined the Christchurch Call to Action in 2019, a global pledge to eliminate terrorist and violent extremist content online. As part of meeting these commitments, Canada has engaged Tech Against Terrorism to develop the Terrorist Content Analytics Platform, which aims to allow smaller technology companies to tackle terrorist content more effectively.[47] It is a centralised platform of verified terrorist content designed to support small technology companies to identify terrorist content, and to help with content moderation decisions.[48] It automates the detection and analysis of verified terrorist content on platforms and is the first and largest dataset of verified terrorist content. Its secondary role is to allow for secure academic research to take place, helping to develop greater understanding of the threat posed by terrorism and extremist content. This will help achieve the commitment under the Christchurch Call to Action to better support small online platforms to build capacity to remove extremist content online.[49]

European Union

The EU currently has a number of legislations that apply to online extremist content. Directive 2017/541 on Combatting Terrorism aims to harmonise member states' legislation on criminalising terrorist offences. More specifically, Article 21 of this directive requires member states to take measures ensuring the fast removal of online content, leaving the type of measure at the member states' discretion.[50] This covers inciting terrorist content, training and recruitment material and other terrorist activities. Some Member States have enacted notice-and-action procedures applicable to online platforms within their national legislation.[51] These include France, Germany and Spain. Under e-Commerce Directive Article 14.3, member states need to establish procedures that govern the removal of or disabling of access to information.[52] Yet, member states retain the authority to interpret this article as they see fit, leading to differences in scope across the EU. The European Commission also agreed in 2016 on a voluntary code of conduct with Microsoft, Twitter, Facebook and YouTube for countering illegal hate speech online.[53]

The EU faces many challenges in classifying and effectively tackling online extremist content. The first and most fundamental is the need for agreement and implementation of common standards and procedures across what will be 27 national legislatures after the Brexit transition period. This challenge at the heart of the EU's working has led to a fragmented landscape among member states.

47 'Press release: Tech Against Terrorism award grant by the Government of Canada to build Terrorist Content Analytics Platform', Tech Against Terrorism, 27 June 2019. Accessed at: 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/.

48 'Update: Initial version of the Terrorist Content Analytics Platform to include far-right terrorist content', Tech Against Terrorism, 2 July 2020. Accessed at: https://www.techagainstterrorism.org/2020/07/02/update-initial-version-of-the-terrorist-content-analytics-platform-to-include-far-right-terrorist-content/.

49 Government of Canada, Public Safety Canada, 'Government of Canada announces initiatives to address violent extremist and terrorist content online', News release, 26 June 2019. Accessed at: https://www.canada.ca/en/public-safety-canada/news/2019/06/government-of-canada-announces-initiatives-to-address-violent-extremist-and-terrorist-content-online.html.

50 European Commission, 'Proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online', 2018/0331 (COD): 3. Accessed at: https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-regulation-640_en.pdf.

51 Ibid.: 122

52 Ibid.: 10

53 United Nations Security Council, Counter-terrorism Committee Executive Directorate, 'More support needed for smaller technology platforms to counter terrorist content', CTED trends alert, November 2018: 4. Accessed at: https://www.un.org/sc/ctc/wp-content/uploads/2019/01/CTED-Trends-Alert-November-2018.pdf.

Figure 11: Existing initiatives in EU Member States on notice and action procedures[54]

MS	Legal act	Legislation in preparation	Illegal content covered
BE	None	Notice & Action (N&A)	
DE	Law on enforcement of rights in social networks (NetzDG)		Hate speech
DK	None		
ES	Royal Decree 1889/2011 on the functioning of the Commission for the protection of IPR – newly modified by Law No. 21/2014.		Copyright infringement
FI	Act No 2002/458 on the Provision of Information Society Services		Copyright infringement
FR	Law No 2004-575 of the 21 June 2004 to support confidence in the digital economy		Only for manifestly illegal content
HU	Act CVIII of 2001 on certain aspects of Electronic Commerce and on Information Society Services		IPR infringement
IT	AGCOM Regulations regarding Online Copyright Enforcement, 680/13/CONS, December 12, 2013		Copyright infringement
LT	Regulation on Denial of Access to Information which was Acquired, Created, Modified or Used Illegally, approved by the Government Resolution No. 881 on August 22, 2007		Horizontal
PL	None	Potentially working on a N&A initiative	
PT	Decree-Law No. 7/2004 of 7 January], Lei do Comércio Electrónico, January 7, 2004		Out-of-court preliminary dispute settlement
SE	Act on Responsibility for Electronic Bulletin Boards		Copyright infringement, racist content
UK	Electronic Commerce Regulations S.I. 2002/2013		Horizontal – it establishes the requirements of a notice

54 Figure taken from European Commission, 2018: 123. Accessed at: https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-swd-408_en.pdf.

For example, with regard to notice-and-action procedures for illegal content being hosted on online platforms, there are no existing rules at the EU level. Instead, only a few member states have introduced regulatory frameworks on notice-and-action procedures. Among these, some did so in the spirit of the e-Commerce Directive, such as France, the UK and Hungary, whereas other operated through a self-standing legal instrument, such as Spain.[55] In some countries, actions to measure, block, filter and remove online content are not in line with Article 10 of the European Court of Human Rights, according to which restrictions on freedom of speech must be 'legal, legitimate and necessary'.[56]

In 2018, the European Commission proposed legislation for extremist content to be removed within one hour of it being uploaded. The legislation would also place a duty of care responsibility on the platforms towards their users.[57] The proposed regulation would also oblige member states to ensure that their authorities and law enforcement agencies have the required capacity to tackle online terrorist content.[58] However, this proved unpopular with certain member states and European Members of Parliament alike. The legislation used a definition of terrorism that was too broad and the one-hour removal requirement was seen as too limiting, potentially creating a culture of censorship.[59] The legislation was amended in the European Parliament to reflect its critics' main concerns.[60] Negotiations around the EU's online terrorist content regulation between the Council, the Commission and the Parliament were placed on hold due to the coronavirus pandemic. In July 2020, the European Commission announced non-binding guidelines under the Audiovisual Media Services Directive, originally adopted in 2018. The guidelines mean that online platforms must ensure their users are protected against hate speech and minors are protected from harmful content.[61] The EU is currently also working on a Digital Services Act, which will seek to 'regulate the online ecosystem across a range of areas including … offensive content'.[62]

France

Officers from the National Police with the responsibility of fighting digital crimes are responsible for any enforcement of legislation. Equally, the Central Office for Combating Information and Communication Technology is responsible for checking whether platforms previously blocked for hosting extremist content still have this content.[63] There is also the L'Office central de lutte

55 European Commission, 2018: 122.
56 Council of Europe, 'Comparative study on blocking, filtering and take-down of illegal internet content', 20 December 2015. Accessed at: https://edoc.coe.int/en/internet/7289-pdf-comparative-study-on-blocking-filtering-and-take-down-of-illegal-internet-content-.html#.
57 European Commission, 2018: 3.
58 Ibid.: 4.
59 Faiza Patel, 'EU 'Terrorist Content' Proposal Sets Dire Example for Free Speech Online', Just Security, 5 March 2019. Accessed at: https://www.justsecurity.org/62857/eu-terrorist-content-proposal-sets-dire-free-speech-online/.
60 Al Cuddy, 'EU struggles over law to tackle spread of terror online', BBC News, 17 April 2019. Accessed at: https://www.bbc.co.uk/news/world-europe-47962394.
61 'Facebook, YouTube, Twitter to face same EU rules on hateful content as broadcasters', EURACTIF, 3 July 2020. Accessed at: https://www.euractiv.com/section/digital/news/facebook-youtube-twitter-to-face-same-eu-rules-on-hateful-content-as-broadcasters/.
62 Samuel Stolton, 'Platform clamp down on hate speech in run up to Digital Services Act', EURACTIF, 23 June 2020. Accessed at: https://www.euractiv.com/section/digital/news/platforms-clamp-down-on-hate-speech-in-run-up-to-digital-services-act/.
63 European Commission, 2018: 117.

contre la criminalité liée aux technologies de l'information et de la communication, essentially the office for cybercrime, the body to which illegal content is reported.

In France, Article 6 of Law no. 2004-575 of 21 June 2004 sets out the liability for hosting platforms. The law states that while companies are not liable for activities or information posted on their platform unless they have knowledge of illegal content and do not take it down, they are, however, responsible for acting on notifications of illegal content. Platforms are required to have a reporting mechanism that allows anyone to report content. If reported, the illegal content must be deleted within 24 hours, or the authorities may report the electronic address of content to the platform, which must block access immediately.[64] Following the terrorist attack on *Charlie Hebdo*'s offices, France also introduced a new act in February 2015 that grants powers to the National Police to take down without a court order any websites containing illegal content.[65] A subsequent degree, 2015-253 of March 2015 applies specifically to direct provocation and/or incitement to terrorism as well as glorification of terrorism.[66]

More recently, France has passed legislation that forces tech companies to take down extremist content within an hour of receiving an order from the French police or face fines of up to 4% of global revenue. The French regulator, the Superior Council of the Audiovisual, will have the power to impose fines on tech companies who breach this legislation. A key challenge for some platforms is the re-uploading of previously detected and removed content. Yet in France the trend set by the Supreme Court appears to be that there is no obligation for platforms to prevent the reappearance of content that has been previously removed – thereby implying a limited scope of duty of care for social media companies.[67] The notice-and-action procedures are limited to illegal content, which may include extremist content, but may not capture all of it.

In particular there seem to be concerns by free speech groups over the consequences of having to take down content. There are two sets of challenges: the first is that this is a challenging requirement for smaller tech companies, who do not have the resources to monitor large volumes of content around the clock.[68] In order to comply with the legislation, they may have to resort to censorship rather than risk facing fines. The second is a concern that the legislation may be used to censor political activism. This in particular demonstrates the difficulty of defining what qualifies as extremist content, as the boundaries are often highly fluid.[69]

64 'Online terrorist propaganda: France and UK put internet giants in the cross-hairs', Jones Day, July 2017. Accessed at: https://www.jonesday.com/en/insights/2017/07/online-terrorist-propaganda-france-and-uk-put-internet-giants-in-the-cross-hairs.
65 European Commission, 2018: 117.
66 Ibid.
67 Ibid.: 10.
68 'France gives online firms one hour to pull 'terrorist' content', BBC News, 14 May 2020. Accessed at: https://www.bbc.co.uk/news/technology-52664609.
69 Benedict Wilkinson and Armida van Rij, 'An analysis of the Commission for Countering Extremism's call for evidence: Report 1 – Public understanding of extremism', Policy Institute, King's College London, 9 December 2019.

Ghana

Ghana has an explicit branch of law enforcement dedicated to cybercrime with input from Europol, Interpol and ISPs.[70] Based on publicly available information, however, it is not clear what Ghana's counterterrorism efforts are. It is also not clear whether there are any ongoing efforts to combat online terrorist content, and if so, what challenges and debates this has brought up.

Japan

The National Centre of Incident Readiness and Strategy for Cybersecurity was established in 2015 and obliges infrastructure companies, such as those of utilities (gas, water, electricity), transport networks and financial institutions to enhance cybersecurity measures proactively.

In 2017, Japan passed a controversial new bill, which targeted conspiracies to commit terrorism and other serious crimes, listing 277 different crimes, including copying music and mushroom-picking in conservation forests.[71] The new laws were criticised for their infringement on civil liberties and vague application. The leaders' statement from the G20 Osaka summit also strongly urged tech companies not to allow the misuse of their platforms for terrorist purposes.[72]

New Zealand

The overarching response to countering terrorism in New Zealand involves coordination between several different government departments, communities and private sector organisations. High-level governance is provided through the Cabinet External Relations and Security committee and the Security and Intelligence Board. New Zealand's overarching strategy is outlined in their Counter-Terrorism Strategy plan, released in February 2020.[73]

The Digital Safety group at the Department of International Affairs is responsible for regulating online content, such as films, videos and other publications that may be classified as existing to 'cause harm'. Harm is defined as any online content that 'describes, depicts, expresses, or otherwise deals with matters such as sex, horror, crime, cruelty, or violence in such a manner that the availability of the publication is likely to be injurious to the public good'.[74] The identification of this material appears to be largely reliant on external reporting systems or internal content classifier systems

70 Kristina Cole et al., 'Cybersecurity in Africa: An Assessment' https://www.researchgate.net/profile/Seymour_Goodman/publication/267971678_Cybersecurity_in_Africa_An_Assessment/links/54e93dca0cf25ba91c7ef580/Cybersecurity-in-Africa-An-Assessment.pdf.

71 'Japan passes controversial anti-terror conspiracy law', BBC News, 15 June 2017. Accessed at: https://www.bbc.co.uk/news/technology-52664609; Robin Harding, 'Japan passes pre-emptive anti-terrorism law', The Financial Times, 15 June 2017. Accessed at: https://www.ft.com/content/75130598-5181-11e7-bfb8-997009366969.

72 Government of Japan, 'G20 Osaka Leaders' statement on preventing exploitation of the internet for terrorism and violent extremism conducive to terrorism (VECT)'. Accessed at: https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/en/documents/final_g20_statement_on_preventing_terrorist_and_vect.html.

73 Government of New Zealand, Officials' Committee for Domestic and External Security Coordinatoin, Counter-Terrorism Coordination Committee, 'Countering terrorism and violent extremism national strategy overview', February 2020. Accessed at: https://dpmc.govt.nz/sites/default/files/2020-02/2019-20%20CT%20Strategy-all-final.pdf.

74 Government of New Zealand, Department of Internal Affairs, 'Objectionable and restricted material'. Accessed at: https://www.dia.govt.nz/Censorship-Objectionable-and-Restricted-Material.

that already exist on social media platforms; the government itself does not appear to conduct its own screening process. The enforcement of any identified objectionable content pertaining to New Zealand is guided by the Films, Videos, Publications and Classifications Act 1993, which was updated with an amendment in 2019 following the Christchurch terrorist attack, which were filmed and made available online. The footage of the attack was voluntary removed by ISPs and social media platform providers as soon as they were notified of objectionable online material by the Digital Safety group. Failure to comply with the act may result in prison sentences of up to 14 years and fines of up to $200,000 NZD.

The unprecedented livestreamed murder of 50 people in Christchurch in 2019 launched the Christchurch Call to Action, co-founded by New Zealand and France to 'eliminate terrorist and violent extremist content online'.[75] This brought together 48 countries, including 31 new countries, to overhaul the Global Internet Forum to Counter Terrorism (GIFCT). GIFCT developed a shared crisis response protocol in 2019, which was tested by Google in New Zealand, to enable the coordinated management of the impact online of any extremist attack.[76]

United Kingdom

In the UK, the Home Office is tasked with counterterrorism legislation and policy. With this mandate, the Home Office works closely with the National Security Centre, established in 2016 and part of the Government Communications Headquarters. The Department for Digital, Culture, Media and Sport (DCMS) has a responsibility for maintaining a safe and open internet.[77] The Home Office also works closely with third parties to develop specific technology to help with the prevention and tackling of online violent extremist content. One group of such stakeholders are tech and AI companies who develop the tools to tackle extremist content, such as ASI data science. The second group are the platforms that are abused by the uploading and sharing of illegal content, such as Facebook, Twitter and Microsoft. There are also arms-length bodies, such as the Commission on Countering Extremism, part of the Home Office, and the UK Council for Internet Safety, which seeks to tackle online harms including hate crime and extremism.[78]

The Home Office established the Police Counter-Terrorism Internet Referral Unit (CTIRU), to which anyone can report suspected content. The CTIRU secured the removal of over 300,000 pieces of terrorist content online.[79] In February 2018, the UK government announced the development of new technology that seeks to analyse videos accurately to determine whether they may be IS propaganda.[80] Using machine learning, this technology was

75 See https://www.christchurchcall.com/.
76 GIFCT, Joint Tech Innovation. Accessed at: https://www.gifct.org/joint-tech-innovation/.
77 Houses of Parliament, Clare Lally and Rowena Bermingham, 'Online extremism', UK Parliament POST, 6 May 2020: 3. Accessed at: https://post.parliament.uk/research-briefings/post-pn-0622/.
78 HM Government, Home Office and Department for Digital, Culture, Media & Sport. 'Online harms – White Paper', April 2019: 36. Accessed at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf.
79 HM Government, Home Office, 'The United Kingdom's strategy for countering terrorism', June 2018: 35. Accessed at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/716907/140618_CCS207_CCS0218929798-1_CONTEST_3.0_WEB.pdf.
80 HM Government, Home Office, 'New technology revealved to help fight terrorist content online', press release, 13 February 2018. Accessed at: https://www.gov.uk/government/news/new-technology-revealed-to-help-fight-terrorist-content-online.

designed specifically for smaller tech companies who, unlike larger companies like Facebook and Youtube, did not have the capacity to develop the tools themselves. The content classifier sits in the upload stream of a platform, which means that a video is rejected before it ever reaches the platform. This is a preventative approach, allowing action to be taken before the potentially harmful video makes it online. The home secretary at the time, Amber Rudd, did not rule out legislation in future to oblige companies who lack other effective monitoring resources to use the tool. The UK has also put in place legislation to act on illegal content online through a notice-and-action framework.[81]

In a joint Home Office-DCMS Online Harms White Paper from last year, the government proposed the establishment of an independent regulator to tackle the challenge of identifying responsibilities for online content regulation.[82] The bill would also be a tool to hold platforms to account for any harmful content spread on their websites. However, there are reports that the bill establishing such a regulator will be delayed in passing through Parliament until 2023 or 2024.[83] Instead, the media and telecommunications regulator Ofcom has been given more powers to make tech companies responsible for protecting people from harmful content, including extremist content.[84] In addition, in September 2019 the UK announced that it would fund research into developing technology to detect videos automatically that have been altered to circumvent existing detection methods.[85] The ambition is for this tool to be made freely available to all tech companies.

United States

In the United States, the Bureau of Counterterrorism within the State Department, led by a coordinator for counterterrorism, has the responsibility to 'develop coordinated strategies and approaches to defeat terrorism abroad and [secure] the counterterrorism cooperation of international partners'.[86] The bureau also works with tech companies to improve information sharing.[87] The Department of Homeland Security works closely with the US's allies, as well as organisations such as Tech Against Terrorism, to tackle online extremist content specifically. The USA also works with the Global Counterterrorism Forum, a multilateral forum that focusses on developing a long-term approach to combating the threat posed by terrorism.

The US counterterrorism strategy of 2018 mentions that to 'combat terrorists' influence online' is a priority area and it says that the USA will aim to combat terrorist use of the online space to recruit, fundraise

81 European Commission, 2018: 20.
82 Houses of Parliament, 2020; Home Office and Department for Digital, Culture, Media & Sport , 2019.
83 'Online harms bills: warning over 'unacceptable' delay', BBC News, 29 June 2019. Accessed at: https://www.bbc.co.uk/news/technology-53222665.
84 Ibid.
85 HM Government, Home Office, 'UK to help develop new tech to stop sharing of terrorist content', press release, 24 September 2019. Accessed at: https://www.gov.uk/government/news/uk-to-help-develop-new-tech-to-stop-sharing-of-terrorist-content.
86 US Government, Department of State, Bureau of Counterterrorism. Accessed at: https://www.state.gov/bureaus-offices/under-secretary-for-civilian-security-democracy-and-human-rights/bureau-of-counterterrorism/.
87 US Government, Department of State, 'Country reports on terrorism 2019', Bureau of Counterterrorism, June 2020. Accessed at: https://www.state.gov/wp-content/uploads/2020/06/Country-Reports-on-Terrorism-2019-2.pdf.

and radicalise individuals while working with partners.[88] The USA has three priorities in terms of countering terrorists influence online: 1) to engage with the private sector; 2) to support counter-messaging efforts by tech companies and civil society organisations; and 3) to protect First Amendment rights.[89]

The First Amendment poses a challenge for US policymakers in regulating online extremist content. The context in the US is somewhat different from that in Europe, because so much speech is protected under the First Amendment. Some content that may be considered illegal in France or the UK may well be legal in the USA.[90] This distinction between 'legal-but-harmful' or 'legal-but-offensive' and plainly illegal content makes it difficult for ISPs to take down offensive content. However, Section 230 of the 1996 Communications Decency Act allows tech companies to moderate offensive content that may be legal.[91] Over the decades, Supreme Court decisions have developed greater nuances to this juxtaposition of free speech versus facilitating violence.[92]

UN Counter-Terrorism Committee Executive Directorate

The UN Counter-Terrorism Committee Executive Directorate (CTED) was established by UN Security Council Resolution 1535 (2004) as an expert body in support of the Security Council's Counter-Terrorism Committee (CTC).[93] Its initial aim was to assess UN member states' implementation of Security Council resolutions on counterterrorism and support their efforts through dialogue. The CTED works closely with the Security Council, the big tech companies through the GIFCT and civil society organisations. It established Tech Against Terrorism, a public–private partnership initiative that seeks to 'support the global technology sector in responding to terrorist use of the internet whilst respecting human rights'.[94]

There currently exist several UN Council Resolutions relating to abuse of the internet for terrorist purposes. Security Council resolution 2129 (2013) notes the evolving interrelation between terrorism and ICT, and the use of technologies such as the internet to commit and facilitate terrorist acts, by allowing the incitement, recruitment, fundraising or planning of terrorist acts.[95] This resolution also re-enforces the mandate of CTED. Resolutions 2354 (2017), 2395 (2017) and 2396 (2017) implore member states to cooperate to prevent terrorist organisations from exploiting the internet, and to work with the private sector and civil society to develop effective measures to prevent the

88 US Government, White House, 'National strategy for counterterrorism of the United States of America', October 2018: 22. Accessed at: https://www.whitehouse.gov/wp-content/uploads/2018/10/NSCT.pdf.
89 US Government, Department of Homeland Security, 'Strategic framework for countering terrorism and targeted violence', September 2019: 24. Accessed at: https://www.dhs.gov/sites/default/files/publications/19_0920_plcy_strategic-framework-countering-terrorism-targeted-violence.pdf.
90 Daphne Keller et al., 'Regulating Online Terrorist Content: A Discussion With Stanford CIS Experts About New EU Proposals', SLS Blogs, 25 April 2019. Accessed at: https://law.stanford.edu/2019/04/25/regulating-online-terrorist-content-a-discussion-with-stanford-cis-directors-about-new-eu-proposals/.
91 Ibid.
92 Victoria L. Killion, 'Terrorism, Violent Extremism, and the Internet: Free Speech Considerations', Congressional Research Service, 6 May 2019: i. Accessed at: https://fas.org/sgp/crs/terror/R45713.pdf.
93 Naureen Chowdhury Fink, 'Meeting the Challenge: A Guide to United Nations Counterterrorism Activities', International Peace Institute, 2012: 45. https://www.ipinst.org/wp-content/uploads/publications/ebook_guide_to_un_counterterrorism.pdf.
94 'March 2020 update', Tech Against Terrorism, 3 March 2020. Accessed at: https://www.techagainstterrorism.org/2020/04/03/march-2020-update/.
95 UN, Security Council Counter-terrorism Committee, 'Public-private efforts to address terrorist content online: A year of progress – what's next?', 14 September 2018. Accessed at: https://www.un.org/sc/ctc/news/event/public-private-efforts-address-terrorist-content-online-year-progress-whats-next/.

abuse of the internet for terrorist purposes.[96] In addition, the CTED in collaboration with South Korea, the GIFCT and Tech Against Terrorism also launched its online Knowledge-Sharing Platform in 2017 to promote the sharing of good practice.[97] The Platform aims to support smaller tech companies to monitor online content and counter violent extremism.[98]

In its November 2018 trends alert, CTED acknowledged the challenge smaller platforms and ISPs face in regulating illegal and terrorist content, which the Knowledge-Sharing Platform is intended to help address. Its publications show the discrepancy in approaches taken by UN member states and tech companies, with a wide range of existing practices.[99] These can vary from tech companies self-regulating, in particular the larger ones such as Facebook and Twitter, to countries that have not yet passed regulations for notice-and-action type measures, such as The Netherlands.

Conclusion

In this report we have provided a policy landscape on how nine difference legislatures tackle online terrorist content. Countries share a recognition that misuse of the internet by extremists and terrorists is a problem that needs to be tackled, both at the state level and from within the private sector. All countries we analysed have or are developing policy and tools to tackle the issue. Yet, while all legislatures appear, bar one, publicly to recognise the threat posed by extremists and terrorists using the online space for their own ends, there is less consensus on the stringency of demands a country should place on tech companies to combat this threat. We also see that countries share several challenges, partners and legislation in tackling online extremist content. However, there also exist differences between these nine legislatures in the means they employ to tackle online extremist content, and the considerations given to freedom of speech rights. Multilateral organisations and institutions, such as the EU and CTED, have at times been challenged by their mandate coming from a variety of countries (CTED), or disagreements at member state level on what is required (EU). Collaborations with global initiatives, such as the GIFCT, and utilising tools developed by organisations, such as Tech Against Terrorism, appear popular.

96 Ibid.
97 Ibid.
98 United Nations Security Council, Counter-terrorism Committee Executive Directorate, 2018: 2.
99 See, for example, United Nations Security Council, Counter-terrorism Committee Executive Directorate, 2018.

Global Network
on Extremism & Technology

CONTACT DETAILS

For questions, queries and additional copies of this report, please contact:

ICSR
King's College London
Strand
London WC2R 2LS
United Kingdom

T. **+44 20 7848 2098**
E. **mail@gnet-research.org**

Twitter: **@GNET_research**

Like all other GNET publications, this report can be downloaded free of charge from the GNET website at www.gnet-research.org.