لا إله إلا الله
الله
رسول
محمد
لا إله إلا الله
الله
رسول
محمد

Tackling Terrorist Content on Social Media

June 14, 2018 | Volume 11 | Issue 32
Stuart Macdonald

Online radicalization is widely regarded as one of today's most pressing security challenges. Its importance has been emphasized by core European institutions, including the Council of Europe, the European Commission, EUROPOL, and the Organization for Security and Co-operation in Europe.[1] These warnings have been echoed by other International Governmental Organizations—the UN's *Plan of Action to Prevent Violent Extremism*, for example, states that the "manipulative messages of violent extremists on social media have achieved considerable success in luring people, especially young women and men, into their ranks"[2]—and by national governments. The UK's Home Affairs Committee has described the use of the internet to promote radicalization and terrorism as "one of the greatest threats that countries including the UK face," whilst the 2015 White House Summit—attended by Ministers from nearly 70 countries worldwide—underscored the need to intensify efforts to counter recruitment and radicalization to terrorist violence.[3]

As concerns about online radicalization have intensified, there have been growing calls for social media companies to do more to remove violent extremist content from their platforms as swiftly as possible. In recent months, the European Commission has threatened to legislate should there not be demonstrable progress,[4] whilst the UK Prime Minister has stated that technology companies need "to move further and faster in reducing the time it takes to remove terrorist content online", adding that ultimately such content should be "removed automatically."

Given both the sheer volume of online content (every minute 350,000 tweets are posted, 300 hours of video are uploaded to YouTube, and 510,000 comments are posted to Facebook, as well as a further 293,000 status updates and 136,000 photos) and the variety of offensive material that may be removed (from sexual exploitation and hate speech to spam and bullying), it is unsurprising that each of these platforms already utilizes artificial intelligence (AI) to assist in the enforcement of their rules and standards. For example, Facebook uses several approaches: *image matching*, so that if someone tries to upload a photo or video that matches a photo

or video that has previously been identified as terrorist, they are prevented from doing so; *language understanding*, which analyzes text that has been removed for praising or supporting terrorist organizations in order to develop text-based signals that can go into machine learning algorithms to detect similar future posts; and *identification of terrorist clusters*, using algorithms to work outwards from pages, groups, posts or profiles that have been identified as supporting terrorism, employing signals such as whether an account is friends with a high number of accounts that have been disabled for terrorism. Moreover, in an effort to prevent terrorists from jumping from one platform to another, Facebook, Twitter, YouTube, and Microsoft have created a shared industry database of hashes (i.e. unique digital fingerprints) for content produced by or in support of terrorist organizations. These hashes allow member companies to identify and remove matching content that violates their policies—and sometimes block such content before it is even posted. As of December 2017, the database contained more than 40,000 hashes, and seven further companies had joined the consortium (Ask.fm, Cloudinary, Instagram, Justpaste.it, LinkedIn, Oath, and Snap).

Attempts to block violent extremist content from social media platforms may be understood as analogous to forms of situational crime prevention. The criminological literature on the latter highlights the fact that such initiatives often have displacement effects. So, while policy-makers' focus on the largest social media companies and the most high-profile terrorist groups, particularly the so-called Islamic State (IS), is understandable, efforts to block violent extremist content online must have regard to the full social media ecology—including the following three dimensions.

Attempts to block violent extremist content from social media platforms may be understood as analogous to forms of situational crime prevention.

First, the use of other strategies. It has been found that, to circumvent technologies that automatically block content, IS supporters have used out-links. Conway et al's 2017 study found that 12.5 percent of the 57,574 tweets analyzed contained out-links.[5] The platform with the most out-links from Twitter was YouTube. Interestingly, however, Facebook did not appear in the top ten, whilst the less-known justpaste.it, sendvid.com, and archive.org all featured in the top six.[6] The use of these smaller platforms appears to be an attempt to "exploit an overlapping ecosystem of services," taking advantage of the fact that smaller companies "don't have the scale

or resources to handle the challenge on their own." Justpaste.it, for example, is a free content-sharing service that allows content to be posted within seconds with no registration required. Owned by Mariusz Zurawek, who runs the site out of his home in Poland, the content posted on Justpaste.it began to include IS propaganda in early 2014. Since then, Zurawek has received a large volume of take-down requests from all over the world. This poses challenges in terms of identifying what content is legal and responding to take-down requests in other languages, as well as capacity and resources.

Second, building on the previous point, attention must be paid to other platforms. In the 29 months from August 1 2015, Twitter suspended over 1.2 million accounts for violations related to the promotion of terrorism. During the same period, IS supporters largely moved their community-building activities to other platforms, in particular Telegram. Telegram is a cloud-based instant messaging service, providing optional end-to-end encrypted messaging. Features include: a self-destruct timer that permanently deletes messages and media after they are viewed; group chats, which users can only join when invited to do so; and, channels, which are public and so can be used to broadcast messages to large audiences. By early 2016, Telegram was being used to share content produced by official IS channels, and IS members and supporters were pushing out more than 30,000 Telegram messages each week.[7] These uses of Telegram form part of a wider movement towards the use of more covert methods. As well as Telegram, other encrypted messaging services, including WhatsApp, have been used by jihadists for communication and attack-planning.[8] Websites have also been relocated to the Darknet. Darknet platforms are, by definition, more difficult to police than the surface and deep webs, meaning they have the potential to function as a jihadist "virtual safe-haven."[9]

Third, there are the activities of other violent extremist groups. Today, IS's Twitter activity "has largely been reduced to tactical use of throwaway accounts for distributing links to pro-IS content on other platforms, rather than as a space for public IS support and influencing activity."[10] Moreover, these throwaway accounts are also suspended before they gain much of a following. Yet, supporters of other jihadist groups—such as Hay'at Tahrir al-Sham, Ahrar al-Sham, the Taliban, and al-Shabaab—experience significantly less disruption.

Conway et al found that, whilst more than 30 percent of pro-IS accounts were suspended within 48 hours of their creation, the equivalent figure for these other pro-jihadist accounts was less than one percent. The latter were also able to post six times as many tweets, follow four times as many accounts, and gain 13 times as many followers as the pro-IS accounts.[11]

As well as broadening the focus to other jihadist groups, it is also important to tackle other forms of violent extremism. Extreme right-wing groups, for example, also have a significant online presence. And, while some steps have been taken to disrupt their presence on the surface and deep webs (e.g. Facebook's decision to ban Britain First

from its platform[12]), it appears that these groups are also beginning to migrate to the Darknet.

Tackling online violent extremist content is rightly an important priority for policy-makers; research has found that the internet plays an important facilitative role in contemporary terrorism.[13] Ensuring that the biggest platforms remove such content as swiftly as possible—or block it altogether—is clearly an essential part of an effective strategy. In pursuit of this objective, however, it is important not to lose sight of the consequential displacement effects. Steps must also be taken to prevent terrorist exploitation of these other dimensions of the social media ecology.

Notes

1. Council of Europe (2014). 'Revised EU Strategy for Combating Radicalisation and Recruitment to Terrorism', 9956/14; European Commission (2015). *The European Agenda on Security*, Com 185; EUROPOL (2016). *European Union Terrorism Situation and Trend Report 2016*. Organization for Security and Co-operation in Europe (2014). *Preventing Terrorism and Countering Violent Extremism and Radicalization that Lead to Terrorism: A Community-Policing Approach*, Office for Democratic Institutions and Human Rights.

2. United Nations (2015). Plan of Action to Prevent Violent Extremism: Report of the Secretary General, A/70/674, 24 December 2015.

3. Home Affairs Committee (2016). Radicalisation: the counter-narrative and identifying the tipping point (8th Report of 2016-17) HC 135; [G] White House Summit to Counter Violent Extremism (2015), https://obamawhitehouse.archives.gov/the-press-office/2015/02/18/fact-sheet-white-house-summit-countering-violent-extremism

4. Gibbs, S (2018) 'EU gives Facebook and Google three months to tackle extremist content' The Guardian 1 March 2018 https://www.theguardian.com/technology/2018/mar/01/eu-facebook-google-youtube-twitter-extremist-content

5. Conway M, Khawaja M, Lakhani S, Reffin J, Robertson A & Weir D (2017) Disrupting Daesh: Measuring Takedown of Online Terrorist Material and Its Impacts. Dublin: VOX-Pol Network of Excellence.

6. The other two platforms in the top six were Google Drive and Google Photos.

7. Prucha N (2016) IS and the Jihadist Information Highway – Projecting Influence and Religious Identity via Telegram. *Perspectives on Terrorism*, 10, 48-58.

8. Malik N (2018) Terror in the Dark: How Terrorists Use Encryption, the Darknet, and Cryptocurrencies. London: The Henry Jackson Society http://henryjacksonsociety.org/wp-content/uploads/2018/04/Terror-in-the-Dark.pdf

9. ibid, iv.

10. Conway et al, n vi above, 30.

11. Conway et al, n vi above.

12. Nouri L (2018) 'Britain First & Facebook: Banned but not Solved?' Centre for Analysis of the Radical Right, 8 April 2018 http://www.radicalrightanalysis.com/2018/04/08/britain-first-facebook-banned-but-not-solved/

13. Gill, P, Corner, E, Thornton, A & Conway M (2015) What are the roles of the internet in terrorism? Measuring online behaviours of convicted UK terrorists. Dublin: VOX-Pol Network of Excellence: http://www.voxpol.eu/download/vox-pol_publication/What-are-the-Roles-of-the-Internet-in-Terrorism.pdf

Stuart Macdonald is Professor of Law and Criminology at Swansea University, U.K.

View PDF