PHYS.ORG

How tech companies are successfully disrupting terrorist social media activity

June 26 2018, by Stuart Macdonald

Platforms for radicalisation? Credit: pixabay/7stonesgfx, CC BY

In June 2017, Google, Facebook, Twitter and Microsoft announced the formation of the Global Internet Forum to Counter Terrorism (GIFCT). The aim of this industry-led initiative is to disrupt the terrorist exploitation of its services. Recently, GIFCT members hailed the achievements of its first year of operations. But, while this progress must be acknowledged, significant challenges remain.

Every single minute there are on average 510,000 comments and 136,000 photos shared on Facebook, 350,000 tweets posted on Twitter and 300 hours of video uploaded to YouTube.

Given this, the biggest companies extensively rely on artificial intelligence (AI). Facebook's uses of AI include image matching. This prevents users from uploading a photo or video that matches another photo or video that has previously been identified as terrorist. Similarly, YouTube reported that 98% of the videos that it removes for violent extremism are also flagged by machine learning algorithms.

Progress so far

One difficulty the social media companies face is that, if a terrorist group is blocked from one platform, it might simply move to a different one. In response to this, GIFCT members have created a shared industry database of "hashes". A hash is a unique digital fingerprint that can be used to track digital activity. When pro-terrorist content is removed by one GIFCT member, its hash is shared with the other participating companies to enable them to block the content on their own platforms.

At its recent meeting, the GIFCT announced that to date 88,000 hashes have been added to the database. So the consortium is on track to meet its target of 100,000 hashes by the end of 2018. Especially so, now that another nine companies have joined the consortium, including Instagram, Justpaste.it and LinkedIn.

These efforts have undoubtedly disrupted terrorists' use of social media platforms. For example, in the 23 months since August 1, 2015, Twitter has suspended almost a million accounts for promoting terrorism. In the second half of 2017, YouTube removed 150,000 videos for violent extremism. Nearly half of these were removed within two hours of upload.

Future challenges

Yet much further work remains. In response to the disruption of their use of Twitter, supporters of the so-called Islamic State (IS) have tried to circumvent content blocking technology by what is known as outlinking, using links to other platforms. Interestingly, the sites that are most commonly outlinked to include justpaste.it, sendvid.com and archive.org. This appears to be a deliberate strategy to exploit smaller companies' lack of resources and expertise.

IS supporters have also moved their community-building activities to other platforms, in particular Telegram. Telegram is a cloud-based instant messaging service that provides optional end-to-end encrypted messaging. This encryption stops messages being read by third parties. And it has been used extensively to share content produced by official IS channels.

This forms part of a wider movement towards more covert methods. Other encrypted messaging services, including WhatsApp, have been used by jihadists for communication and attack-planning. Websites have also been relocated to the Darknet. The Darknet is a hidden part of the internet that is anonymous in nature and only accessed using specialist encryption software. A 2018 report warned that Darknet platforms have the potential to function as a jihadist "virtual safe-haven."

In addition, recent research has found that supporters of jihadist groups other than IS experience significantly less disruption on Twitter. Supporters of these other groups were able to post six times as many tweets, follow four times as many accounts and gain 13 times as many followers as pro-IS accounts.

It is also important to respond to other forms of violent extremism. Extreme right-wing groups also have a significant presence on platforms such as YouTube and Facebook. While steps have been taken to disrupt their presence online, such as Facebook's decision to ban Britain First from its platform, it appears that these groups are also beginning to migrate to the Darknet.

Overreach

Just as there is an issue of reaching terrorist social media, there are also challenges relating to potential overreach. Machine learning algorithms cannot be expected to identify terrorist content with 100% accuracy. Some content will be wrongly identified as terrorist and blocked or removed. But the challenges here go further than just applying the threshold correctly. They also concern where the threshold should be drawn in the first place.

The difficulties in defining terrorism are well known. Summed up by the slogan "One person's terrorist is another's freedom fighter", one of the most controversial definitional issues is that of just cause. Should a definition of terrorism exclude those such as pro-democracy activists in a country ruled by an oppressive and tyrannical regime? According to many countries, including the UK, the answer is no. As one Court of Appeal judge put it: "Terrorism is terrorism, whatever the motives of the perpetrators."

If social media companies take a similar approach, this could have some significant ramifications. Indeed, there are already worrying examples. In 2017, thousands of videos documenting atrocities in Syria were removed from YouTube by new technology aimed at extremist propaganda. These videos provided important evidence of human rights violations. Some existed only on YouTube, since not all Syrian activists and media can afford an offline archive. Yet the alternative—to seek to distinguish between just and unjust causes—is fraught with difficulties of its own.

At a time when social media companies face increasing pressure to do more to tackle terrorist exploitation of their platforms, the progress made during the GIFCT's first year is welcome. But it is only the first step.

This article was originally published on The Conversation. Read the original article.

Provided by The Conversation

Citation: How tech companies are successfully disrupting terrorist social media activity (2018, June 26) retrieved 25 April 2024 from https://phys.org/news/2018-06-tech-companies-successfully-disrupting-terrorist.html