15 April 2019

Country: Global

by: Grant Williams

It is no secret that terrorists have used online platforms such as Facebook, YouTube and Twitter as a means to recruit and radicalise people online. As soon as it became clear that groups like the Islamic State were actively using these platforms to recruit followers and spread propaganda, social media companies started to employ artificial intelligence (AI) technology to flag, and remove hateful content before it could spread further. As of March of this year, this software was removing more than one million accounts per day from Facebook alone.

However, in the wake of the Christchurch massacre in New Zealand, it has become clear that AI tools are not picking up on far right, extremist messages in the same way. Some blame social media companies
for employing a
 double standard
in evaluating violent content, but the reality is more complex. White supremacists frequently communicate through coded language, 'in jokes' and sarcasm, circumventing the AI tools meant

to spot only explicitly violent content, allowing an undercurrent of extreme hate to flourish online.

One of these techniques is 'echos,' a simple tactic where a user places multiple brackets around a word to label someone, or something, as Jewish online – (((like this))). The symbol originated in 2014 on an anti-semitic podcast, The Daily Shoah, where echo effects were added when saying Jewish names. This has subsequently trickled down and spread to online platforms such as Twitter.
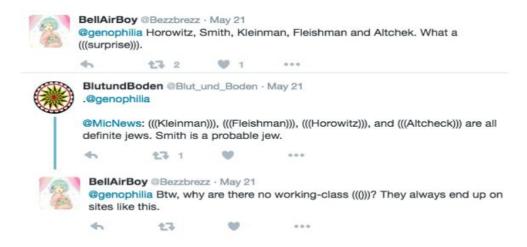

Figure 1: Example of 'echos' being used on twitter

Another common coded message used is a reference to the number 1488. The 14 refers to the total words which make up a popular mantra with white supremacists: "We must secure the existence of our people and a future for white children." The 88 references the eighth letter of the alphabet, H, to imply, 'Heil Hitler.'

Joshua Fisher-Birch, a content review specialist at the Counter Extremism Project, questioned social media platforms' ability to remove the content. "I don't think [social media companies] have the capabilities, even at a basic level, when white supremacist content is flagged to act on it," he told CNN.

Donald Trump recently downplayed the growing threat of white nationalists as, "a small group of people that have very serious problems." However, Trumps statement is contradicted by a recent report, conducted by The Anti-Defamation League (ADL), which showed that white supremacy rallies in the US rose by 20 per cent over the last year, from 76 to 91. It also showed that the distribution of white supremacist propaganda has increased by 182 per cent during the

same period. A 2016 study from George Washington University's Program on Extremism shows that US white supremacist Twitter accounts' followings grew by 600 percent between 2012 and 2016, even though tech companies were actively removing hateful content during that time.

"The internet has played a powerful role in developing the [beliefs] of contemporary racism," writes Andrew Jakubowicz, professor of social and political sciences at the University of Technology in Sydney, Australia. "The political economy of the internet favours freedom over control, facilitated by technologies that magnify the anonymity of racist protagonists."

"It shouldn't surprise us that bigots are early adopters of technology," said Jonathan Greenblatt, the chief executive of the ADL, to the New York Times. "Their noxious views are difficult to circulate openly. They can post something to Twitter or Facebook and achieve exponential reach under a cloak of anonymity."

With Donald Trump playing down concerns over the rise of white supremacy and social media companies not succeeding in removing hateful content, some governments are looking to take drastic action in order to combat it.

One such country is India. Their government has drafted a policy of rules which would enable them to force internet companies to remove content from their sites   .
Prime Minister, Narendra Modi, is trying to impose regulations which would stop Indians from seeing "unlawful information or content". The UK, Australia and Singapore are looking to follow suit.

Some experts, however, believe that this would be used as an attack on the right for free speech. "The government intervention that they propose is potentially more damaging than the problem they want to solve," wrote Niam Yaraghi, professor of Operations and Information Management at University of Connecticut's School of Business.

"If conservatives believe that certain businesses have enough power and influence to infringe on their freedom of speech," he said, "how can they propose government, a much more powerful and influential entity, to enter this space?"

Giving any institute the ability to censor content online inevitably raises concerns as to their motives for removing it. In order to ensure human bias is removed from the equation regarding whether something is deemed as 'hateful', AI seems to be the obvious method. However, as mentioned, that is not without its flaws.

"The main problem," said Pedro Domingos, a professor of computer science at the University of Washington and author of The Master Algorithm, to CNN, " is that the [far-right extremist] content is too variable and multifarious to be reliably distinguished from acceptable content by the filtering algorithms that tech companies use, even state-of-the-art ones."

The Lawyers' Committee for Civil Rights Under Law made a crucial step in the right direction though, in March 2019, when they were victorious in an effort to get Facebook to change their policy and block content that promotes white nationalism and white separatism as acceptable hate speech.

"There is no defensible distinction that can be drawn between white supremacy, white nationalism or white separatism in society today," said Kristen Clarke, president and executive director of the Lawyers' Committee for Civil Rights Under Law, after their victory.

"By maintaining this distinction, Facebook ended up providing violent racists a platform that could be exploited to promote hate. While we are pleased that Facebook is taking long overdue action, we know well that communities are still reeling from the rise in hate and racially motivated violence, and that extensive remedial action must be taken to ensure that hate is eliminated root and branch across the platform."

"It's clear that these concepts are deeply linked to organized hate groups and have no place on our services," said Facebook, via a statement, after agreeing to the policy change.

For more on online hate speech, check out our Get The Trolls Out project. For our coverage of the Christchurch attacks in the media, read [Dr. Verica Rupar's article on the ethics of care](#) , and Madeline Rose Leftwich on the [Islamophobia in the Australian media](#)

. Don't miss our
[*interview with Jean-Paul Marthoz*](#)
on the double standards of Islamist terrorism and far right terrorism.