CURRENT INTELLIGENCE REPORT

Date: December 12, 2019
Subject: Social Media Threat Monitoring to Predict and Prevent Mass Violence

KEY FINDINGS

- 56 percent of active shooters who carried out mass attacks between 2000 and 2013 leaked their intent to commit violence prior to the attack.
- Information relevant to threat monitoring is more likely to be found on obscure, less regulated platforms, rather than mainstream social media platforms.
- People who later commit a violent attack are more likely to use emotionally charged words, more likely to use direct pronouns, and less likely to use words about the external world in their online posts.
- Eight warning behaviours of individuals who could present a concern for lone-actor terrorism are Pathway (research, preparation), Fixation, Identification as an agent of a cause, Novel aggression (a small unrelated act of violence), Energy burst, Leakage, Direct threat, and Last resort (a declaration which indicates increased distress).
- 10 characteristics of individuals who could present a concern for lone-actor terrorism are Personal grievance and moral outrage, Framed by an ideology, Failure to affiliate with an extremist group, Dependence on the virtual community, Thwarting of occupational goals, Failure of sexual-pair bonding (evidence of failure to form lasting intimate relationships), Changes in thinking and emotion, History of mental disorder, Creativity and innovation (in regards to tactical planning of an attack), and History of criminal violence.
- Possible limitations with automated threat monitoring tools include bias, human-computer interaction issues and accountability.
- Misinterpretation of non-verbal communication, foreign languages, slang, non-familiar cultural references, and non-standard English dialects are potential limitations of automated monitoring tools.
- Recommendations for human expertise to support an automatic threat monitoring system include multilingual analysts and linguists; investigative experience; ability to access critical data and resources; and knowledge of privacy laws, copyright acts and violations of social media platforms' terms of service.

Background

This report aims to document views of risk professionals, research organizations, psychologists and journalists on the use of social media monitoring to predict and prevent mass violence. The report will highlight social media platforms that have been identified as being useful for threat monitoring, including

obscure, anonymous forums. Key problems and limitations encountered with automated monitoring will be identified and assessed, as well as the challenges of human-computer interaction.

Following two mass shootings over one weekend in August 2019 in Dayton, Ohio, and El Paso, Texas, the director of the FBI issued an order for agents to conduct online threat assessments in an effort to prevent similar attacks[1]. In many incidents of lone shootings and terror attacks, online threats of violence have preceded the incident. By the end of August 2019, over 20 people in the United States had been arrested for making threats of violence online. One of these was 18-year-old Justin Olsen, who had threatened to carry out shootings at Planned Parenthood locations on meme-sharing site iFunny, under the username "ArmyOfChrist"[2] [3]. According to an FBI spokesperson, the posts had first been flagged by an FBI office in Anchorage, Alaska, and agents continued to monitor Olsen's posts. A search of Olsen's house two days after his arrest located a collection of weapons including 15 rifles including AR-15 style rifles, 10 semi-automatic pistols, over 10,000 rounds of ammunition, and a machete in the teenager's car.

Objectives of the Report

In the aftermath of two August 2019 mass shootings in Ohio, and Texas, USA, and two consecutive March 2019 mosque shootings in Christchurch, New Zealand, social media threat monitoring has been the subject of much discussion. The objective of this report is to document views on what is needed from technology and human expertise to utilize threat monitoring of social media platforms.

Social media platforms

According to a report by Smart Insights in February 2019, there are just over 3.4 billion social media users worldwide in 2019[4]. When thinking about social media platforms, mainstream sites usually come to mind, such as Facebook, Instagram and Twitter, however information relevant to threat monitoring is more likely to be found on obscure, less regulated and anonymous platforms[5] [6]. Facebook and Instagram are limited by privacy laws from being accessed by data companies seeking to conduct threat intelligence[7].

The following are some examples of more obscure social media platforms:

---

[1] CNN August 22, 2019 https://www.cnn.com/2019/08/21/us/mass-shooting-threats-tuesday/index.html?utm_medium=social&utm_term=link&utm_content=2019-08-22T03%3A06%3A05&utm_source=twCNN

[2] Daily Best, August 13, 2019 https://www.thedailybeast.com/fbi-justin-olsen-arrested-for-threatening-massacre-had-10000-rounds-of-ammo

[3] Embedded Podcast "This is not a joke", November 7, 2019 https://www.npr.org/podcasts/510311/embedded

[4] Smart Insights https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research

[5] Echosec https://www.echosec.net/blog/social-media-threat-intelligence-are-facebook-and-instagram-relevant

[6] Kroll https://www.kroll.com/en/insights/publications/cyber/social-media-threat-monitoring-preempt-violence

[7] Echosec https://www.echosec.net/blog/social-media-threat-intelligence-are-facebook-and-instagram-relevant

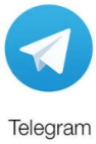
- **8kun** – the forum formerly known as "8chan" was relaunched on November 2, 2019, under the name "8kun"[8]. The website consists of user-created message boards. 8chan had been taken down after the man accused of killing 50 people at two mosques in Christchurch, New Zealand, posted a manifesto to the site[9]. The perpetrator of a shooting which killed 20 people at a Walmart in El Paso, Texas, had also posted to the site 20 minutes prior to the incident. The website has been filtered out from Google Search as a result of the alleged presence of child pornography[10], and the site is currently only reliably accessible on the dark web[11].

- **Reddit** – an online forum and news aggregator with over 300 million users. Reddit users have personal accounts, usually with anonymous usernames.

- **Raddle.me** – an online forum, with approximately 600,00 active users, which is notorious for its shoplifting message boards.

- **Discord** – a VoIP (voice over IP) application and messaging program with over 40 million active users. The site is popular with gamers. It also became popular with alt-right users and has been associated with cyber bullying, organizing alt-right rallies, adult content, and discussion of illegal activity.

- **Telegram** – an instant messaging and VoIP service. Telegram has been known to contain discussions regarding illegal activities around the world.



[8] Business Insider November 3, 2019 https://www.businessinsider.com/8kun-8chan-website-known-for-shooting-associations-relaunched-as-2019-11
[9] Business Insider March 25, 2019 https://www.businessinsider.com/new-zealand-bans-christchurch-shooter-manifesto-livestream-2019-3
[10] Wikipedia https://en.wikipedia.org/wiki/8chan
[11] Slate November 11, 2019 https://slate.com/technology/2019/11/8chan-8kun-white-supremacists-telegram-discord-facebook.html

- **VK** – a Russian-based social media platform, also known as "Vkontakte". The network has almost 100 million active users worldwide.

- **iFunny** – an image-based picture sharing application. The site is popular for sharing memes, but has become associated among its popularity with white nationalist users.

Automated tools

There is increasing pressure for law enforcement to work with social media companies to improve online monitoring to prevent mass attacks. The FBI announced in July 2019 that it aimed to establish an *"early alerting tool"* to scrape social media sites for threats of mass violence[12]. The FBI stated that it would require the tool to capture the following for persons of interest: social networks, usernames, emails, IP addresses, telephone numbers, real-time alerts about content based on keywords, geolocation, and photo-tagging.

Another example of automated social media monitoring tools already in use are those provided by private risk organizations to other companies. The organizations offer monitoring of a brand's social presence as well as 24/7 automated scanning over the internet and dark web of social media, message boards and public forums for high-risk content, with automatic removal provided. Services also include scanning for accounts impersonating the businesses. Results are analysed by specialists, reported, and then clients are assisted with their threat response, including removing problems or reporting to authorities[13].

An organization specializing in threat intelligence in the United States, Digital Stakeout, provides examples of the type of automated searches available[14]:

- **RSS Monitoring** - monitoring from any valid RSS (Really Simple Syndication) feed – a regularly updated XML file containing a list of newly published content from a website
- **Location monitoring** – monitoring location-based social media to identify the risks related to specific geographies and places
- **Web page monitoring** – monitoring content on specific URLs
- **Search-engine monitoring** – monitoring for information on top search engines
- **Dark web monitoring** – monitoring dark web forums and marketplaces for threats

[12] U.S. Department of Justice, June 2018 https://www.fbi.gov/file-repository/pre-attack-behaviors-of-active-shooters-in-us-2000-2013.pdf
[13] Threat Minder https://www.threatminder.com/services
[14] Digital Stakeout https://www.digitalstakeout.com/platform/scout/monitors/risk-reconnaissance

- **Page monitoring** – monitoring of specific social media pages

Warnings and indicators

According to the article *"How to tell when social media posts signal a mass shooter in the making"* by the director of the Center for Terrorism and Security Studies in Massachusetts, USA, which refers to a study by the director of the Center for Terrorism and Security Studies and a professor of psychology, perpetrators of mass attacks rarely make clear public declarations of their intentions[15]. However, they might leave signals that, if located and interpreted correctly, could offer the opportunity for law enforcement to prevent attacks. Automated tools would need to be programmed to identify these signals and distinguish between posts by people who are simply venting frustrations online and those that intend to carry out violence.

The U.S. Department of Justice released a report in June 2018 that identified behaviours exhibited by people prior to carrying out mass attacks between 2000 and 2013 (including periods prior to the launch of many popular social media platforms)[16]. The report included details on types of "leakage" of intent to commit a violent act in more that half of the cases studied.

Key findings from the report included the following[17]:

- 77 percent of active shooters spent a week or longer planning their attack
- 56 percent of active shooters leaked intent to commit violence prior to the attack
- 88 percent of the active shooters aged 17 and younger leaked intent to commit violence, compared with 51 percent of adult active shooters who leaked their intent
- In cases where active shooters had pre-planned targets for their attack, over half made threats prior to the attack. However, in 65 percent of those cases, no threats were communicated towards a specific target

The article *"How to tell when social media posts signal a mass shooter in the making"* compared the language of online postings of people who had allegedly committed a mass attack with posts from people who had expressed ideological beliefs online, but when investigated by law enforcement were found to have no plans of violence. The article reported the following findings[18]:

- People who later became violent were more likely to use emotionally charged words such as *"shit," "hate", "hurt", "stab", "murder",* etc.

[15] Fast Company October 27, 2019 https://www.fastcompany.com/90422442/how-to-tell-when-social-media-posts-signal-a-mass-shooter-in-the-making
[16] U.S. Department of Justice, June 2018 https://www.fbi.gov/file-repository/pre-attack-behaviors-of-active-shooters-in-us-2000-2013.pdf
[17] U.S. Department of Justice, June 2018 https://www.fbi.gov/file-repository/pre-attack-behaviors-of-active-shooters-in-us-2000-2013.pdf
[18] Fast Company October 27, 2019 https://www.fastcompany.com/90422442/how-to-tell-when-social-media-posts-signal-a-mass-shooter-in-the-making

- People who later became violent were less likely to use words about the external world, such as *"people," "world," "state"* and *"time"*
- People who later became violent were more likely to use direct pronouns such as *"you", "they", "me"* in their posts

According to the report *"The Concept of Leakage in Threat Assessment"* there are a number of motivations for a violent individual sharing their intentions, including[19]:

- A desire to create fear and intimidation associated with the impending attack
- A need to seek attention for themselves
- An inability to contain emotions associated with planning the attack
- A desire for the leakage to be memorialized after their death or after the event, and to gain notoriety for themselves

A report published in 2016 *"The Clinical Threat Assessment of the Lone-Actor Terrorist"* by forensic psychologist J. Reid Meloy and research assistant Jacqueline Genzman, went into more depth by listing 18 indicators (warning behaviors and characteristics) of individuals who present a concern for lone-actor terrorism[20]. This research was conducted in the context of assessments by mental health professionals.

The indicators were provided as follows.

Warning Behaviors:

- **Pathway** – indicators are research, planning and preparation (for example, researching and purchasing weapons)
- **Fixation** – *"an increasingly pathologic preoccupation with a person or a cause, accompanied by a deterioration in social and occupational life".* Indicators could be researching extreme materials and expressing radical beliefs online
- **Identification** – identifying with previous attackers or assassins; identifying themselves as an agent of a cause (for example Justin Olsen AKA "ArmyOfChrist"); closely associating with military, and law enforcement paraphernalia
- **Novel aggression** – a small unrelated act of violence believed to be a way to test the individual's ability to carry out violence
- **Energy burst** – an increase in activity connected to the target or to preparation of the attack, for example increased online searches connected to the individual's beliefs, visits to shooting ranges, visits to the target location. Or an increase in activity to "tie up" lose ends, such as meeting with close family
- **Leakage** – communication to a third party of an intent to do harm to a target - through letters, diaries, journals, blogs, videos on the internet, emails, voice mails and social media

[19] J. Reid Meloy, Mary Ellen O'Toole, 2011, "The Concept of Leakage in Threat Assessment" http://drreidmeloy.com/wp-content/uploads/2015/12/2011_theconceptofleakage.pdf

[20] J. Reid Meloy PhD, Jacqueline Genzman, BA, 2016 http://drreidmeloy.com/wp-content/uploads/2015/12/Meloy-and-Genzman-online-August-2016.pdf

- **Direct threat** – the communication of a direct threat to the target or to law enforcement
- **Last resort** – a declaration in words or actions which indicates increased distress or desperation

Characteristics:

- **Personal grievance and moral outrage** – personal grievance could include a major loss in relationships or employment. Moral outrage could include the identification with a group that has suffered
- **Framed by an ideology** – in incidents of a terrorist attack, this would mean the presence of a religious belief system, political philosophy, or one-issue conflict that could be used to justify the act
- **Failure to affiliate with an extremist group** – a lone-actor terrorist being rejected by a group with which they had wanted to affiliate
- **Dependence on the virtual community** – evidence of the use of the Internet communication through social media, chat rooms, emails, etc. in relation to extreme views or planning attacks
- **Thwarting of occupational goals** – *"a major setback or failure in a planned academic and/or occupational life course"*
- **Failure of sexual-pair bonding** – evidence of failure to form lasting intimate relationships
- **Changes in thinking and emotion** – expression of views becoming more imposing on others
- **History of mental disorder**
- **Creativity and innovation** – evidence of creativity and innovation in regards to tactical planning of an attack
- **History of criminal violence**

Human Expertise

On August 23, 2019, the Global Association of Risk Professionals (GARP) in New Jersey, USA, wrote that an automatic monitoring system will only be as effective as the human expertise (specialists, public safety officials, security professionals) that support it. The organization published the following recommendations for human expertise[21]:

- Investigative experience that translates into knowing where trends, patterns and shifts are developing
- Experience and insight to avoid pitfalls, such as those associated with "profiling" or collecting data from incorrect stolen identities
- Ability to access critical data and resources, such as global law enforcement agencies
- Knowledge and experience of privacy laws, copyright acts and violations of social media platforms' terms of service
- Multilingual analysts and linguists to aid in translating messages

[21] Global Association of Risk Professionals (GARP), August 23, 2019, https://www.garp.org/#!/risk-intelligence/technology/data/a1Z1W000003m8rbUAA

Limitations

Automated tools have obvious limitations when it comes to interpreting findings and human expertise is needed to work alongside these tools. Rachel Levinson-Waldman, senior counsel to the Brennan Center for Justice's Liberty and National Security Program in the USA raised some skepticism on the ability for automated tools to detect threats, *"Even the best machine learning tools in the private sector are, at best, 70% to 80% effective at identifying threat indicators. In cases where private companies use algorithms to detect threats, the worst possible consequence is having an account terminated. When you talk about law enforcement getting into the game, when they have the power of investigation and prosecution...the stakes are incredibly high"*[22].

The Partnership on AI, an organization in California which studies best practices in artificial intelligence technologies, published the "Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System" in 2019. The report highlights the following possible limitations:

- Bias – predicted probabilities by automated tools could be either too high or too low for specific groups of people based on their race, gender, age or socioeconomic class, for example[23].
- Human-computer interaction issues – humans must not over-rely on the accuracy of automated systems. It must be clear to all users (including judges and lawyers, among others) how the data was captured and the predictions made are uncertain
- Accountability – users must ensure that regular evaluation, monitoring, and auditing of tools is carried out

The Brennan Centre for Justice released a report on May 22, 2019, which assessed the United States' Department of Homeland Security's (DHS) collection and use of information located from social media profiles. The information is used to evaluate the potential security risks posed by national or foreign travelers. The report raised difficulties that had been faced by DHS staff when interpreting their findings. For example, in 2012 a British citizen was denied entry at a Los Angeles airport after DHS agents located and misinterpreted a Twitter post by the individual stating he would *"destroy America"*, which he had intended as slang for partying, and another that he would *"dig up Marilyn Monroe's grave"* - a reference to a TV show.

The Brennan Centre report also raised the problem that interpretation is even harder when the language used is not English and the cultural context is unfamiliar, pointing out that *"If the State Department's current*

[22] WIJA, August 6, 2019 https://wjla.com/news/nation-world/fbi-seeks-tools-to-monitor-social-media-detect-mass-shooters-before-they-strike

[23] Partnership on AI, 2019, Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system

plans to undertake social media screening for 15 million travelers are implemented, government agencies will have to be able to understand the languages (more than 7,000) and cultural norms of 193 countries"[24].

The article *"Racial Disparity in Natural Language Processing: A Case Study of Social Media African-American English"* by Su Lin Blodgett and Brendan O'Connor also highlights the problems of interpreting the language of English speakers using non-standard dialects or slang. The article points out that this language can be incorrectly flagged as foreign by automated tools, providing the example of one post "Bored af den my phone finna die!!!!" (which can be loosely translated as "I'm bored as f*ck and then my phone is going to die") which had been identified by an automated tool as Danish with 99.9 percent confidence[25].

Interpretation is also a problem when it comes to non-verbal communication on social media. For example, if an individual "loves" an article on Facebook regarding a terror attack, should this be interpreted as them signalling support for the act, or sending love to people affected by the attack?

Summary

There is no single warning behaviour, indicator or algorithm for successfully identifying an individual who is planning mass violence. However, social media monitoring can provide critical information and this tool is best leveraged as one of many diverse strategies alongside highly skilled multilingual analysts with investigative experience[26].

Automated social media threat monitoring has its limitations and must not be relied upon alone to make decisions and to detain or continue the detention of persons of interest. When using information gathered to make decisions, it must be made clear when this information is presented how it was captured and the uncertainty behind the predictions being made. Users must also ensure that the tools receive independent review by third parties, as well as regular evaluation, monitoring, and auditing of these tools is carried out[27].

[24] Brennan Centre for Justice, May 22, 2019, https://www.brennancenter.org/our-work/research-reports/social-media-monitoring

[25] Su Lin Blodgett and Brendan O'Connor, "Racial Disparity in Natural Language Processing: A Case Study of Social Media African-American English", 2017, https://arxiv.org/pdf/1707.00061.pdf

[26] Global Association of Risk Professionals (GARP), August 23, 2019, https://www.garp.org/#!/risk-intelligence/technology/data/a1Z1W000003m8rbUAA

[27] Partnership on AI, 2019, Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system