# RSIS WEBINAR SERIES ON "DRUMS: DISTORTIONS, RUMOURS, UNTRUTHS MISINFORMATION & SMEARS"

Event Report

1st, 2nd and 4th December 2020

RSiS S. RAJARATNAM SCHOOL OF INTERNATIONAL STUDIES
Nanyang Technological University, Singapore

NSCS NATIONAL SECURITY COORDINATION SECRETARIAT

# RSIS WEBINAR SERIES ON "DRUMS: DISTORTIONS, RUMOURS, UNTRUTHS, MISINFORMATION & SMEARS"

## Event Report

**1st, 2nd and 4th December 2020**

**Report on the Workshop organised by:**
Centre of Excellence for National Security (CENS), S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University (NTU), Singapore

**Supported by:**
National Security Coordination Secretariat (NSCS), Prime Minister's Office (PMO), Singapore

**Rapporteurs:**
Jennifer Yang Hui, Dymples Leong Suying and Eugene EG Tan

**Editor:**
Muhammad Faizal Bin Abdul Rahman

The panel sessions of the workshop are captured in the conference report with speakers identified. Q&A discussions are incorporated without attribution.

# Table of Contents

# Executive Summary

The Centre of Excellence for National Security (CENS) at the S. Rajaratnam School of International Studies in Nanyang Technological University (NTU) organised their fourth annual workshop on "DRUMS: Distortions, Rumours, Untruths, Misinformation & Smears" as a three-part webinar series on 1st, 2nd and 4th December 2020.

The webinar series explored trends and issues in information manipulation – disinformation, misinformation and online falsehoods – that bedevilled the world in 2020. These trends and issues affect national security by influencing the processes and outcomes of elections and political discourse, public health and safety, and trust in media sources and national institutions. Expert speakers at the webinar series also examined governments' and civil society's efforts, and digital campaigns to counter the information disorder.

The first webinar explored how malicious domestic and foreign geopolitical actors used information manipulation to undermine elections integrity and social peace, and influence online political discourse in western democracies and Asia.

On elections integrity, democratic societies need a comprehensive framework to protect themselves against risks from hybrid campaigns. These campaigns comprise cyber-attacks that target elections' digital infrastructure, and disinformation targeting political figures and election results to sow public confusion and disorder. Disinformation could cause harm even when cybersecurity measures prevent hostile actors from attacking the voting systems. Regarding the Twitter-space, Western experts' analysis unveiled a network of inauthentic accounts that had behaved in a coordinated manner to target western democracies. This campaign supports the Chinese Communist Party (CCP) narratives about issues such as COVID-19 that could be critical of China.

During the Myanmar 2020 elections, political parties, elements linked to the military, and extremist movements used disinformation tactics such as hoaxes and fake media outlets to target political opponents and minority groups. Throughout the U.S. 2020 elections, the coordinated use of peer-to-peer texting and influencers to engage voters presented regulatory loopholes and new vulnerabilities. Hostile domestic and foreign actors could exploit them to influence voting decisions and undermine trust in the elections system.

The second webinar explored how the COVID-19 infodemic aggravates the impact of the pandemic. It complicates the health crisis, fuelling racism and xenophobia, and could persist as countries seek the supply and distribution of safe and affordable vaccines.

In Malaysia, studies unveiled that COVID-19 misinformation and disinformation incorporates images and videos to appear credible, and spread mostly via the encrypted messaging platform WhatsApp besides Facebook. Misinformation and disinformation promotes scams, ill-will towards different races, religions and nations; and feeds religious apprehension that could drive anti-vaccine ideas. In Indonesia, civil society organisations such as MAFINDO help counter misinformation and disinformation resulting from the politicisation of the pandemic. Conspiracy theories and clickbait tactics exploit people's concern for their families and religious identities to spread distrust and drive political contestations. Regarding healthcare organisations, the World Health Organisation (WHO) and healthcare officials worldwide are the targets of information attacks. For example, online memes, slurs and hashtags that are racist and conflate with anti-China messages had targeted the Director-

General of the WHO each time he tweeted. In Europe, lesser-known social media platforms enable an ecosystem of online sub-communities where COVID-19 conspiracy theories and extremist ideas thrive. Non-governmental organisations such as EU DisinfoLab play essential roles in fact-checking and raising the public's media literacy against the infodemic.

The third webinar explored how journalists and media outlets, deepfake technology, social media influencers and digital advertisements play a crucial role in the information environment. They could influence opinions and spread harmful content across all demographics of internet users.

On digital advertising, the work of grassroots social media campaign Sleeping Giants examined ad revenue models. The campaign reached out to brands that had inadvertently placed their advertisements on the extremist Breitbart website which circulates online disinformation. In preventing harmful content, digital marketers and brands should be of aware where their ads appear. On social media influencers and internet celebrities, societies should not underestimate their influence on topical conversations. Influencers use tactics such as "insta-vagueing" and memes. They are adept in

leveraging social media platforms to optimise online engagements and disseminate messages. Deepfakes are growing in popularity among celebrities and politicians who use the technology for outreach and entertainment purposes. Experts are concerned that exposure to deepfakes could foment the liar's dividend, and studies unveiled that a third of Singaporeans and Americans surveyed had unknowingly shared deepfakes. On influence operations, the media could be a conduit for it depending on the state-backed and private interests that fund them. Fake news outlets could conceal their origins to evade foreign influence countermeasures and engage journalists to reach their target audiences.

# Webinar One: Elections, Misinformation and Disinformation 1st December 2020

## Managing Comprehensive Elections Risk

Liisa Past, Head of Cyber Security Business Development, Cybernetica and former Chief National Cyber Risk Officer at the National Security and Defence Coordination Unit, Estonia

**Summary:** Democratic societies need a comprehensive framework to mitigate risks from hybrid campaigns that seek to undermine elections integrity. This framework focuses on protecting the act of voting from any form of coercion or falsification. Firstly, risk management should ensure that the technological infrastructure, which societies use for elections are secure. Secondly, there is a need to mitigate risks that emanate from the information space, but they may be the hardest to counter during elections. Overall, the framework should aim to ensure a safe and fair election.

- Risks to elections can arise from opportunistic political actors and various adversaries who have more persistent goals. Persistent adversaries are typically patient and well-resourced. Their long

term goal is to undermine democracy in the targeted state by de-legitimising the electoral process. There are no limits in the use of tools for attack and selecting their targets. Hence, persistent adversaries are likely to mount integrated operations to achieve their goals. For example, Kyiv's power supply disruptions during Christmas and winter in 2015 and 2016 were a display of political power by Russian-linked actors. The incident also demonstrated how cyber-attacks could threaten a society's way of life.

- The cyberspace is increasingly becoming a domain for states to project their political power through influence or coercion. In this domain, cyber-attacks have evolved from being small threats to societies in the 1990s to becoming an existential threat to democracies today. Cyber-attacks can vary in tactics: (i) shutting down all the information and communication technology (ICT) systems in a state (e.g. Estonia in 2007), (ii) targeting a specific critical infrastructure (e.g. Iran in 2010) or a particular company in the private sector (e.g. Sony Pictures in 2014), and (iii) conducting an online disinformation campaign to discredit a national leader (e.g. Georgia in 2008).

- Societies that embrace digitalisation would also digitalise their democracy. The process of implementing technology has implications for the security of the electoral process. Cyber-attacks can compromise the electoral process by breaching the ICT infrastructure and auxiliary systems that states use to cast and count votes. These cyber-attacks can breach ICT systems by weaponising vulnerabilities such as the backdoors found in Switzerland's online voting system. Hackers could have used the backdoors to remain undetected when altering legitimate votes. Estonia discovered and corrected a fundamental cryptographic flaw in its digital national I.D. card system. There could be questions on the election results' integrity if governments did not detect and patch these vulnerabilities.

- Disinformation can undermine confidence in elections integrity through attempts to falsify or misinterpret the election results. During the Ukrainian elections in 2014, the Russian Channel One T.V. News had portrayed a far-right extremist candidate Dmytro Yarosh as winning 37 per cent of the votes. In reality, Dmytro Yarosh had a low chance of winning with 0.7 per cent of

the votes. During the eve of the Latvian elections in 2018, the popular social networking website Draugiem.lv was reportedly hacked to display pro-Russia messages. These incidents demonstrated how disinformation could cause public confusion and panic, especially during tense elections. Therefore, elections officials must provide accurate and precise information to the public to mitigate any disinformation attempts to undermine election integrity.

- In sum, a comprehensive election risk management framework has four parts in the form of concentric circles. At the centre of the framework is ensuring people's confidence in the voting process. Secondly, the technology that governments develop specifically for the casting and counting of votes must be secure. Thirdly, auxiliary systems, facilitators and vendors who support the election process must be secured. Finally, integrated information operations require both the protection of ICT systems and the information space to ensure the integrity of conversations that are important to democracy.

## Pro-Chinese Communist Party (CCP) Twitter Networks

Raymond M. Serrato, independent open source investigator, anti-disinformation specialist and former social media analyst at the United Nations High Commission for Human Rights

**Summary:** Quantitative studies of the twitter-space unveiled that the CCP has been using Twitter to influence online discussions on various issues that may be critical of China. These issues include the origins of the COVID-19 virus and the series of protests in the United States (U.S.). The studies also examined how many inauthentic accounts with dubious names had supported pro-China viewpoints on specific issues.

- Coordinated inauthentic behaviour on social media entails a network of inauthentic accounts working together to mislead people. At the core of this behaviour is online platforms amplifying or disseminating content on specific issues. Inauthentic accounts make regular posts and participate in a disinformation campaign that a malicious actor(s) orchestrate(s). Monitoring this campaign is challenging as controversial issues also generate high volumes of social media posts from authentic accounts. Genuine political

activity online often appears coordinated too. However, the timing of the posts can unveil whether they are from real persons or indicative of coordinated inauthentic behaviour.

- Analysis of Twitter account characteristics can indicate whether suspicious activities are happening on social media. These characteristics include the number of digits in account handles, the use of same social media profiles across various accounts, whether the accounts' naming conventions suggest computer generation and the creation of accounts occurred in batches during specific times. Observable patterns of coordinated behaviour include high rates of activity mainly accounts re-tweeting each other's tweets, making numerous tweets on particular issues and tagging other social media users. More sophisticated patterns include co-tweeting, re-tweeting, mentioning other accounts' user names and using multiple accounts to tweet similar content around the same time.

- The study of pro-CCP Twitter networks comprises the retrieval of 828,646 followers and 179,112 friends from two CCP official accounts at the start of the COVID-19 pandemic. A sample of data from 5,156 accounts and 445,570 tweets

was extracted and processed using the R software. The analysis revealed that 27 per cent of the followers were created in 2020, particularly between March and April. This account creation pattern coincided with a key CCP official promoting a conspiracy theory on the origin of the virus and U.S. President Trump describing COVID-19 as the “China virus”. A high number of accounts contain naming conventions that use eight digits or more, which is an indicator of possible inauthentic accounts. The tweet timeline revealed that replies constituted nearly 65 per cent of tweets, and most of them were directed at President Trump, especially after he began using the term "China virus". Narrative warfare was happening online between pro-CCP Twitter accounts on the one hand, and President Trump and other accounts that use the term "China virus" on the other hand.

- The analysis of the narrative timeline using Chinese and English keywords across different topics revealed that COVID-19 was the primary theme in the Twitter accounts' tweets. Other themes include the protests in the U.S, Hong Kong and Taiwan. Tweets on U.S. protests, for example, integrated content relating to #blacklivesmatters and #votetrumpout2020. The

tweets had also highlighted social and political tensions in the U.S. possibly influencing the public order situation there and pointing out American hypocrisy. For example, a tweet that was critical of the U.S. had mentioned #FiveDemandsNotOneless, which is the same hashtag that the Hong Kong protestors use. This tactic attempts to contrast the protests and police response in the U.S. with those happening in Hong Kong.

- Analysis of indicators of suspicious activities on social media can help in the identification and examination of possible coordinated inauthentic behaviour. However, this analysis may not be sufficient for unveiling information operations by hostile state actors. Indicators used to assess "authenticity" and "coordination" are heuristics. Publicly available data have limitations and may be insufficient for determining agents' identity or threat actors – unless they are careless – responsible for the disinformation efforts. Coordination patterns alone are inadequate for detecting disinformation campaigns. Nonetheless, a contextual understanding of issues can help in investigating the possible motivations of the hostile actors.

## The Myanmar 2020 Elections: The Role of Fake News and Disinformation

Hunter Marston, PhD candidate in International Relations at the Australian National University in the Coral Bell School of Asia-Pacific Affairs and independent consultant with GlobalWonks LLC

**Summary:** Both misinformation and disinformation were present during the campaigning in the lead up to the Myanmar 2020 elections. Misinformation originated from various sources, including rumours circulating on social media, word of mouth, and print media. Disinformation was more insidious and the work of political actors, social media accounts with links to the military, and influencers who spread falsehoods that target political opponents and minority groups.

- Myanmar faces a high risk of exposure to fake news and disinformation. It is a young democracy undergoing a rapid pace of internet penetration and intense social and political conflicts internally. The country has seen the internet penetration rate grow from less than one per cent in 2012 to almost half of the population in 2018. Mobile phone penetration rate had increased from less than two per cent in 2010 to over 90 per cent in 2017. Facebook is a popular source of

information for people in Myanmar, and up to 73 per cent of the people there rely on social media for news. The number of Facebook users has grown exponentially in Myanmar from about eight million users in 2015 to 26.3 million in 2020, with most users aged 13 to 34.

- Various domestic actors such as the military, extremist movements such as Ma Ba Tha movement and elements that support the Union Solidarity and Development Party (USDP) have used disinformation campaigns. The disinformation targets were the ruling party National League for Democracy (NLD), Aung San Suu Kyi (ASSK), Rohingya/Muslims and USDP. The USDP, in particular, has close ties with the military, which previously ruled the country for decades. The disinformation campaigns, among other things, portrayed the NLD as being weak in protecting Buddhism in Myanmar, which is predominantly Buddhist. In October, a month before the elections, about 50 per cent of fake news specifically targeted ASSK compared to 10 per cent in July. Closer to elections, fake news became more personalised in discrediting individual politicians.

- Military-linked accounts had shared content that supports USDP and is critical of ASSK. In October 2018, the New York Times reported that many of these accounts displayed inauthentic behaviour by posing as celebrities and influencers to spread false narratives targeting the Rohingya/Muslims and ASSK. After that, Facebook announced that it had removed a number of these accounts for coordinated inauthentic behaviour. These accounts were influential as they had almost 12 million followers. This development showed that social media accounts that spread disinformation have a vast following in the country. However, military-linked accounts were not solely responsible for spreading disinformation. ASSK's Facebook account, for example, had claimed that military atrocities against the Rohingya never happened.

- Besides social media accounts, the disinformation campaigns also used websites such as Radio Free Myanmar (RFM), which mimic U.S-based media outlet Radio Free Asia to present itself as a legitimate news source. The RFM started in 2019, but it began posting content – mostly disinformation – only in the months leading up to the elections. RFM used innovative tactics to evade scrutiny, including posting

screenshots and other images, making it more difficult for algorithms to detect. Its followers were encouraged to copy and paste pictures instead of forwarding when sharing content. Also, RFM has a worldpress.com site to appear legitimate and uses fake profiles for its news authors.

- Given the enormity of the disinformation problem in Myanmar, the civil society and young people are playing a significant role in exposing fake news, online hate speech and disinformation, and raising the level of digital and media literacy. Civil society organisations that work to mitigate social media risks include Panzagar, Myanmar ICT for Development Organisation (MIDO), Phandeeyar and the Myanmar Tech Accountability Network (MTAN). Panzagar, for example, ran an anti-hate speech sticker campaign in 2014. The sticker was downloaded 2.7 million times and used in 12.9 million Facebook posts. MIDO has a fact-checking initiative that challenges RFM. It also reports on fake propaganda emails that originate from accounts claiming to be journalists and news organisations or mimicking government institutions.

## Subversive Avenues of Influence: The U.S. 2020 Elections and the Role of Peer-to-Peer Text Messaging and Social Media Influencers

Katie Joseff, Senior Research Associate at the Centre for Media Engagement, The University of Texas at Austin.

**Summary:** The COVID-19 pandemic has necessitated new election campaigning modes to replace in-person canvassing of votes and fundraising. These new modes of campaigning include peer-to-peer text messaging and social media influencers to reach out to more people who are spending more time online. However, these new modes also give rise to grey areas that election campaign laws may not regulate and hence present opportunities for foreign influence and foreign funding involvement. It is difficult to assess these new modes' reach and impact, which operate in the spectrum between grassroots mobilisation and covert manipulation.

- The U.S. 2020 elections happened amid increased fear, distrust, and division, which were exacerbated by the pandemic. The run-up to the elections had seen a rise in protests relating to police brutality, resistance to pandemic control

measures and increased support for QAnon conspiracy theory. These issues have occurred amid COVID-19 when Americans are spending more time online. Social media platforms were better prepared this time to counter disinformation and misinformation. They have shut down several foreign influence operations and implemented policies early to regulate political advertisements in a non-partisan way. They also added “friction” into the system to encourage users to fact-check before sharing messages. Their efforts had seen different levels of success and detractors accusing their policies as inconsistent or politically biased.

- Peer-to-peer texting (SMS or MMS) as a means of election campaigning increased significantly in swing states. Voters received personalised text messages instead of broadcast messages by political action committees (PACs). For example, one text message claimed to be from a “Democratic volunteer with APP PAC” and said that Joe Biden had endorsed sex change operations for children. A regulatory loophole allowed parties who use personalised text message to avoid disclosing the sender or funder of the message and seeking the consent of the person receiving the text. Personalised text

messages also appear to garner a higher response rate than emails or other means of electronic campaigning. One reason for the higher response rate is relational organising, which is a friend-to-friend outreach. It leverages high trust relationships between campaign volunteers who send the message and their friends and other contacts who receive the message.

- The level of political mobilisation of social media influencers was higher in the U.S. 2020 elections. Although the presidential candidates already have large numbers of followers on their social media accounts, political campaigns increased social media nano-influencers' mobilisation to reach out to voters. Nano-influencers may have fewer than 10,000 followers, but they elicit more trust from followers who they engage more frequently at a more personal level. It is less expensive to mobilise nano-influencers as being an influencer is not their primary occupation. Still, they may be influential as respected community figures such as pastors and mummy bloggers with targeted audiences.

- The mobilisation of social media influencers has its risks. Firstly, the coordinated use of

influencers to manipulate messaging is an ethically grey area. It can mimic a grassroots movement and hence appear harmless. Secondly, unlike political advertising, no laws require influencers to disclose whether they are paid and their funding source. The lack of law presents a regulatory loophole for foreign influence, dark money and domestic influence operations. Thirdly, influencer mobilisation's distributed nature makes it difficult to identify signs of coordination and whether the influencers received funding from the same sources. Social media companies cannot track off-platform payments that influencers receive.

The panel sessions of the workshop are captured in the conference report with speakers identified. Q&A discussions are incorporated without attribution.

# Webinar Two: Health Crises, Misinformation and Disinformation 2nd December 2020

## COVID-19 False Information in Malaysia

Harris Zainul, Analyst at Institute of Strategic and International Studies (ISIS) Malaysia and Minister-appointed member of the National Youth Consultative Council of Malaysia

**Summary:** A study by ISIS Malaysia determined that COVID-19-related misinformation comes in a wide variety of narratives and generate varying levels of harm. There is a need for appropriate policy responses to address the different possible motivations for creating and disseminating misinformation during the pandemic. Anti-vaccine propaganda is an area that the government should address as it is becoming more organised and influential and threatens to derail the fight against the pandemic.

- A study of articles published on Sebenarnya.my, the Malaysian government's fact-checking platform, showed that most pandemic-related misinformation had spread via encrypted instant messaging app WhatsApp, followed by

Facebook. Unlike the "open" forms of social media like Facebook, detecting false information on WhatsApp is challenging due to the platform's end-to-end encryption. Identifying falsehoods that spread via WhatsApp is therefore limited to user reporting or tip-offs. The majority of the false content was in textual format. Of these, 18% were a combination of texts and images. Misinformation that uses visual forms such as photos and videos can appear more credible in taking incidents and claims out of context to push a narrative.

- The narratives in misinformation mostly contain false claims relating to actions that the authorities had taken (52%) and community spread (16.2%). The Malaysian government's assistance plans to provide cash and handouts to the needy became the subject of misinformation. Claims of COVID-19 cases in various parts of the country also abound. The former creates confusion among citizens on how they should adhere to appropriate standard operating procedures (SOPs) in uncertain times. The latter, on the other hand, add panic to an already stressful situation. Pandemic-related misinformation, therefore, reduces public trust and worsen citizens' uncertainty about the situation.

- Falsehoods also fuel feelings of ill-will towards people from other countries, of different ethnicity and religion. For example, misinformation regarding migrant workers and refugees had caused a lack of concern towards the turning away of 200 Rohingya refugees on a boat in April 2020. People may become desensitised to the plight of migrants and refugees.

- Pandemic-related scams also use falsehoods to exploit people's fears and deceive potential victims. For instance, scammers pretended to be government officials and claimed that they could help victims process cash handout applications. Another tactic that scammers use is the creation of a fake app for the Employees Provident Fund.

- There is a need for different policy responses to address different types of falsehoods that cause different types of harm. The study determined that "Troll, Provoke or Genuine Belief" are the most likely motivations for sharing pandemic-related misinformation in Malaysia.

- Heavy-handed responses against the creators and disseminators of falsehoods may be ineffective in curbing the spread and appear disproportionate. People who share

misinformation out of ignorance, for example, should not be subjected to prosecution. In addressing the trust deficit that resulted from pandemic-related misinformation, stakeholders should be more involved in disseminating facts. Authorities should improve the clarity and coherence of their communication and policies.

- Anti-vaccination propaganda threatens to undermine Malaysia's efforts to fight the pandemic. Science denialism does not have much traction in Malaysia, given that most people wear face masks and adhere to social distancing. There are a few instances of conspiracy theories relating to virus transmission as compared to the U.S. A YouGov survey found that 82% of Malaysians would accept vaccination. However, experts have warned that anti-vaccine activists are becoming more organised, vocal and influential. These activists exploit religious concerns regarding vaccines' permissibility in general for Muslims and misconceptions of vaccine side effects. Going forward, persuading people vulnerable to vaccine-related misinformation is the next big challenge and a matter of life and death.

## MAFINDO's role in Fact-Checking and Educating COVID-19 Awareness during the infodemic

Aribowo Sasmito, Co-Founder and Head of Fact-Checker Committee, Indonesia Anti-Slander Society (MAFINDO)

**Summary:** Misinformation, especially during a pandemic, is a severe threat to public health and promotes distrust among Indonesian citizens. The resultant infodemic worsens when there is a politicisation of the pandemic by domestic actors.[1] Civil society has a significant role in tackling the infodemic online through crowdsourcing to debunk falsehoods and offline through public education.

- Like other countries, Indonesia is grappling with the infodemic that poses challenges to efforts to fight the pandemic. Coronavirus-related misinformation has swept extensively across the

[1] The World Health Organisation (WHO) describes an infodemic as "…excessive amount of information about a problem, which makes it difficult to identify a solution. They can spread misinformation, disinformation and rumours during a health emergency." See Department of Global Communications. "UN tackles 'infodemic' of misinformation and cybercrime in COVID-19 crisis," *United Nations*, March 31, 2020. https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-and-cybercrime-covid-19.

country and online, making it difficult for people to make accurate judgments and identify solutions. Online rumours abound even before the announcement of the first two confirmed COVID-19 cases in March. Claims of confirmed cases in hospitals across Indonesia's many provinces and cities went unabated, coupled with advice for people to take precautions by wearing masks. Most misinformation also promotes unproven remedies such as gargling with warm salt warm water to neutralise the coronavirus.

- Conspiracy theories promote distrust among Indonesians. For example, the “Plandemic” video, which promotes falsehoods about the COVID-19 pandemic, had circulated widely in Indonesia. There have also been calls to reject the coronavirus vaccine based on the misconception that it contains a microchip that can remotely kill Muslims. Such conspiracy theories appear believable because they appeal to the Indonesians’ cultural values that emphasise religious identity and family well-being.

- Most falsehoods take content out of context and use clickbait tactics to grab attention. For example, MAFINDO's reverse image search

showed that a photo with the caption "Dead bodies on the street of Wuhan" is a picture of an art installation in Germany. It is also easy for falsehoods to manipulate visual medium such as photos and videos to appear compelling. For example, images claiming that Chinese people are susceptible to the coronavirus due to their uncooked meat consumption, and pictures showing alleged mass graves of Italian COVID-19 casualties had circulated in Indonesia widely.

- Disinformation resulting from the politicisation of the pandemic aggravates the infodemic in Indonesia. As political contest persists between the government and the opposition, the pandemic is weaponised for subversive messaging. For example, the opposition had seized missteps in crisis communications by senior political leaders who claimed that the coronavirus could not survive in Indonesia's hot weather. Within their echo chambers, the political supporters amplify such falsehoods on social media. They also use false narratives that can trigger fear and anger in disinformation, calling for people to defy large-scale social distancing measures (PSBB) that the government has implemented.

- MAFINDO's strategy to tackle the infodemic is a combination of offline and online measures. As part of their public awareness campaigns, MAFINDO volunteers collaborate with organisations such as the Indonesian National Police to distribute educational pamphlets, vitamins, and masks. They also conduct talk shows on radio and television to educate the public about misinformation. MAFINDO has established a task force to prioritise and coordinate COVID-19-related fact-checking. Online, MAFINDO crowdsources for misinformation that requires fact-checking. Articles that MAFINDO has fact-checked are available on its various social media pages. MAFINDO also has a WhatsApp chatbot that connects with its database and enables people to use their mobile phones to submit articles for fact-checking.

## COVID-19 misinformation targeting World Health Organisation & other conspiracies

Edward Tian, computer science and journalism student at Princeton University and contributor at Bellingcat

**Summary:** The World Health Organisation (WHO), public health officials worldwide and their COVID-19 related messages have been the target of online attacks. These attacks undermine trust in the WHO and public health officials and drown the messages they disseminate to fight the pandemic. Open-source investigations have proven to be crucial in detecting coordinated inauthentic behaviour driving such attacks. Policymakers in public healthcare must understand these attacks' nature to better prepare for a future health crisis.

- Disinformation disrupts COVID-19 crisis communication and undermines trust in international and national healthcare organisations. For example, the Director-General of the World Health Organisation (WHO), Dr Tedros Adhanom Ghebreyesus, was the target of racist memes and online slurs whenever he tweets. These online attacks began in April 2020 and peaked in June/July 2020 before decreasing.

- Open-source investigations revealed that coordinated inauthentic behaviour drives these online attacks. These attacks use memes and hashtags #ChinaLiedPeopleDied and #TedrosLiedPeopleDied and alleged that China and the WHO had misrepresented and

downplayed the pandemic. Most Twitter accounts that conducted these attacks have recent creation dates and connect with a network of similar accounts. Other suspicious indicators include activity levels only during critical events such as election periods, and accounts that are "sleepless spammers" tweeting most of the time and resting for two hours at most. Twitter may be aware of these accounts' inauthentic behaviour as it has removed some of them for violating community guidelines.

- Open-source investigations can unravel the nature of online attacks against public health officials during a pandemic. The Twitter API is an excellent tool to collect data that investigators can then analyse using software such as Python. Twitter API, however, may be limited if investigators try to survey more than 6000 tweets. Twitter has introduced initiatives that allow investigators and researchers to purchase more data at affordable rates. Nonetheless, it is a good investment for investigators to procure quality data scraping tools.

- Furthermore, open-source investigations can help researchers examine other disinformation campaigns such as far-right conspiracy theories

that the QAnon movement is spreading. The U.S. Federal Bureau of Investigations (FBI) has designated this movement as a domestic terror threat. This movement has become more influential and disseminated various pandemic-related disinformation (e.g. hydroxychloroquine being a cure for COVID-19). Investigations suggest that the movement may have a social media "war room" that guides meme warfare to its supporters.

- In ensuring effective crisis communication, public health officials need to work with researchers and invest in open-source investigations to understand online attacks' nature. Although healthcare organisations do not seek to counter information operations such as engaging trolls, they need to stay ahead of malicious actors who weaponise the internet and social media. These actors use disinformation to challenge the scientifically-backed information that is central to public health education.

## The Rise of Covid-19 Conspiracy Theories in Europe

Roman Adamczyk, Research Coordinator for EU DisinfoLab

**Summary:** The COVID-19 pandemic has created a climate of fear and uncertainty that allows the rise of conspiracy theories. Online sub-communities are increasingly interacting in an ecosystem of lesser-known social media platforms where conspiracy theories thrive. Censorship has its limitations as believers of conspiracy theories can manipulate it to justify falsehoods. The best way to counter conspiracy theories is by engaging the civil society to debunk them and restore trust in national institutions fighting the pandemic.

- Conspiracy theories flourish in the climate of fear and uncertainty that the COVID-19 pandemic created. At the start of the pandemic, people had questions regarding the coronavirus's nature and how long lockdowns and other containment measures would be in force. Malicious actors took advantage of the situation to spread inflammatory messages against the government and mainstream media. Claims that the coronavirus is human-made and is a tool for secret agendas sought to discredit minority groups, the government and foreign powers. Conspiracy theories on vaccines and other purported cures misled people and added to the uncertainty.

- Conspiratorial actors such as alternative doctors, anti-vaccination activists and extremist groups have been peddling conspiracy theories for many years. Large Facebook groups and fringe websites focusing on the COVID-19 pandemic are also online echo chambers where members affirm each other's' belief in such theories and reject evidence to the contrary. Some public figures, such as celebrities and politicians, endorse conspiracy theories that online sub-communities share online. In the U.K., attacks against 5G masts happened after some celebrities pushed conspiracy theories claiming that the technology spread the coronavirus. Public figures generate widespread impact when they endorse falsehoods due to their large following on social media and greater access to traditional media.

- Demonstrations against COVID-19 measures in Europe indicate that conspiracy theories and extremist ideas have gained traction. The globalisation of narratives that are anti-experts, anti-elite and anti-government is happening and framing COVID-19 measures as repression. American anti-vaccine activist Robert R. Kennedy Jr., for example, gained fame in Europe after conducting demonstrations in Berlin.

Extremist groups with international appeal are exploiting socio-political tensions arising from the pandemic. For instance, Far-Right groups had attacked the German Parliament amid the pandemic. The former English Defence League leader Tommy Robinson in the U.K. had violated COVID-19 measures by participating in a rally.

- The fight against conspiracy theorists and extremists remains challenging due to the vastness of the online ecosystem and profits in peddling falsehoods. Famous conspiracy theorists were able to draw their believers away from major social media platforms to smaller ones such as the Twitter-like platform “Parler” that claims to be committed to free speech, and the video site “Odysee”. Disinformation on popular platforms like YouTube has also been overlooked relative to Facebook and Twitter. An example of conspiracy theorists profiting from disinformation is the French pseudoscience propaganda film “Hold-Up”, which earned more than €182,000 through crowdfunding. The film also garnered more than 4 million views on YouTube.

- Censorship can backfire if believers of conspiracy theories manipulate its narratives to

justify falsehoods. Moreover, societies in the U.S., Europe and South Korea are seeing a deficit of trust in the media. Some people believe the media has exaggerated the impact of the pandemic and spread misleading information. Therefore, the next best defence against conspiracy theories is civil society. It is important not to understate civil society efforts to support fact-checking, conduct media literacy programmes and fight distrust.

The panel sessions of the workshop are captured in the conference report with speakers identified. Q&A discussions are incorporated without attribution.

# Webinar Three: Technology and Journalism – Countering Misinformation/Disinformation 3rd December 2020

## How Disinformation Works

Nandini Jammi, Brand Safety Advocate and Co-Founder of Check My Ads

**Summary:** Breitbart gained prominence during the 2016 U.S. election and served as a conduit to create and circulate disinformation online.[2] News about the website trended into mainstream media. Companies that were unaware of their ad placements on websites such as Breitbart removed their ads after Sleeping Giants reached out to them on social media. Companies should improve awareness of the sites that their ads appear.

- Research into the ad revenue model of Breitbart revealed that famous brands had their advertisements placed on the website due to the

[2] Breitbart "…publishes racist and incendiary speech against Muslims, Jews, LGBT people, immigrants, women and …" See Jammi, Nandini. "It's About Free Speech, Says Tech CEO Cashing In On Breitbart Ads," *Medium*, January 25, 2017, https://nandoodles.medium.com/its-about-free-speech-says-tech-ceo-cashing-in-on-breitbart-ads-cd60329284b.

way algorithms in programmatic advertising work. Sleeping Giants, a grassroots social media campaign that the speaker co-started then reached out to these brands. The campaign sought to educate brands about programmatic advertising in which Google Ads and Facebook Ads automate advertisement placement across the internet. Companies that were previously unaware of their ad placements on websites such as Breitbart removed their advertisements after Sleeping Giants reached out to them on social media. Consequently, Breitbart saw its ad revenue declined by 90% in three months, lost about 4,000 advertisers and was dropped by over 30 ad networks.

- Brand safety technology facilitates keeping ads away from unsafe websites such as those that promote hate and violence. The technology, which uses keyword block lists and semantic intelligence algorithms, may also help keep ads away from controversial or sensitive issues. Blocklists are not a 100 per cent effective tool as the algorithms underpinning it scans only the URLs of articles and cannot identify the context in which banned keywords are used. For example, some keyword block lists banned the word “coronavirus” after scanning URLs of

articles, resulting in news sites and media outlets losing ad monetisation revenue.

- Similarly, semantic intelligence algorithms have their limitations. The algorithms can analyse articles on a website by using natural language programming to ascertain content and sentiments. However, the algorithms are unable to analyse context accurately. Analysis of context is a task that requires human involvement, but this is lacking in automated processes that rely heavily on algorithms. Hence, malicious actors can earn ad revenue to fund their disinformation campaigns through programmatic advertising and avoid banned keywords and sentiments.

- In the digital advertising industry, there is currently a lack of education and awareness among marketers on where their ads are appearing and how they can keep their brands safe from being unwittingly associated with hate speech and disinformation. The implications can be harmful to brand reputation and entrench bigotry, hate, racism and disinformation that undermine societies. Ultimately, companies should be aware and accountable for the online content that their ads are funding and be more judicious in their ad placements.

## Influencers, Misinformation and ‘Under the Radar’ Strategies

Dr Crystal Abidin, Senior Research Fellow & ARC DECRA Fellow in Internet Studies at Curtin University

**Summary:** An internet celebrity is an individual who is native to the internet, has high visibility and receives views and acknowledgement from an audience online. An influencer possesses an internet celebrity's attributes, gains fame from positivity and skills, and intentionally monetises viewer engagements. It is critical not to underestimate social media influencers' role in their circuits of (mis)information and innovative “under the radar” communication strategies.

- Influencers are proficient in utilising communication strategies that are prevalent across the internet and savvy in disseminating messages while cutting through white noise on social media. One of the strategies is “insta-vagueing”, which uses ambiguous quotes that contain specific messages. This strategy creates more impressions and encourages followers to be curious about the influencer's real message. This engagement strategy can also create opportunities for the influencers' social media

account to trend in search engine optimisation indexes and could even appear in mainstream media and news.

- The relationship between influencers and algorithmic manipulation is complicated. Influencers tend to bandwagon major social media platforms such as YouTube in which algorithms are a significant factor in a post's reach. Hence, algorithms can amplify influencers' content that may polarise sentiments and promote click baits across social media channels and the political spectrum.

- Influencers are also savvy in using memes on social media groups (e.g. Facebook, YouTube and Instagram) and instant messaging channels (e.g. WhatsApp and Telegram). Memes that contain misinformation or taken out of context may then spread to family social media groups and WhatsApp family chat groups. Memes that originate from meme factories are significant as they are the product of coordinated online action and can drive new perceptions and social norms. Such memes use humour to set boundaries, shape public conversations and provide alternative ways to discuss specific topics.

Memes can normalise conversations on taboo topics by breaking the stigma of discussing them.

- In sum, it is essential not to underestimate social media influencers' role and the use of memes in their circuits of (mis)information. The impact of their messaging is neither marginal nor inconsequential. Their persuasive powers also come from their ability to use shared lingo and internet vernacular in reaching out extensively to people of various demographic groups including Generation Z, millennials and senior citizens.

## Trust, Disrupted: Examining Deepfakes on Social Media

Asst Prof Saifuddin Ahmed, Wee Kim Wee School of Communication and Information, Nanyang Technological University

**Summary:** Deepfakes, digitally-manipulated videos of people such as politicians and celebrities, are becoming more viral on social media. Its use ranges from education to entertainment purposes. Due to the realism heuristic, people tend to trust visuals more than text. Deepfakes can give credence to liars to deny the authenticity of factual content, hence

creating a less truthful version of facts.[3] The potentially dangerous implications of deepfakes to democracy is a cause for concern.

- Research that the speaker conducted indicates that one in two people in Singapore is unaware of deepfakes. In the research's survey, respondents who consume news and are more interested in politics tend to be more aware of it. Respondents surveyed in the U.S. indicated that they were more concerned over deepfakes than those surveyed in Singapore. A significant cognitive bias is the "third-person perception" in which people perceive that media messages, including deepfakes, would have more significant effects on other people but not on themselves. These people feel that they are less vulnerable to deepfakes and rated their ability to recognise deepfakes higher than others.

- 33.2% of respondents in Singapore and 39.1% of respondents in the U.S. indicated that they had

[3] Liar's dividend can be defined as a trend in which people use deepfakes to delegitimise reliable information. See Ahmed, Saifuddin. "Who inadvertently shares deepfakes? Analysing the role of political interest, cognitive ability, and social network size," *Telematics and Informatics* (2020): 101508, p.2. https://www3.ntu.edu.sg/CorpComms2/Research%20Papers/deepfake.pdf.

shared a deepfake by accident which they later found out to be a hoax. Among these respondents, demographics (e.g. age, gender) were not factors in sharing deepfakes. Research suggests that frequent exposure to deepfakes can potentially result in the increased sharing of deepfakes. People who are more politically interested are also more likely to share deepfakes. Furthermore, people who have the lower cognitive ability (i.e. general intelligence) are more likely to share deepfakes than those with higher cognitive ability.

- Most people are vulnerable to deception by deepfakes, although their biases and perceptions make them believe the contrary. Exposure to deepfakes can deceive people and foment scepticism of news on social media. When this scepticism happens, people have the subjective feeling of alienation and distrust towards news on social media in general. People with medium and low cognitive abilities, who are concerned about deepfakes and had accidentally shared them on social media, may become less confident about their ability to distinguish deepfakes from authentic videos. Over time, the heightened distrust and scepticism may also make people more sceptical of accurate and factual news.

- There should be more research to understand public engagement with deepfakes, looking ahead. There is a need for more public education about deepfakes. Public education can help to inoculate social media users to the potential harms of deepfakes, and several tools can facilitate it. For example, gamification tools such as "Bad News" can improve social media users' confidence and cognitive ability to identify fake news, including deepfakes. More importantly, more efforts should continue and enhance digital literacy, particularly for people who are less able to distinguish facts from falsehoods.

## Influence Operations and the Media

Alicia Wanless, Director, Partnership for Countering Influence Operations at Carnegie Endowment Centre for International Peace

**Summary:** Influence operations can co-opt all forms of communications. During the American Civil War, newspapers were the medium of influence operations. Influence operations can be either positive or negative and what distinguishes between the two is the values underlying the operations. For example, many democracies believe it is acceptable

for them to promote their governance model to other countries. In distinguishing between acceptable and unacceptable influence operations, more objective criteria (e.g. transparency of origins, quality of content, content intention, and scale of operation) can be useful.

- The media can intentionally and unintentionally become a conduit for influence operations. For example, the media can serve as a means for information laundering, particularly for information obtained through hack-and-leak. The media can be used to obfuscate the origins and build up the credibility of the story. This approach can work well in the lead up to elections when political candidates are quick to pick up and use any information which discredits their opponents. The media may also espouse a particular agenda based on the intentions of their financial backers. Besides state backing, private interests (e.g. billionaire Robert Mercer funded Breitbart, which actively campaigned for Donald Trump in the 2016 U.S. election) can influence the media's position.

- Influence operations can also leverage the media to hire journalists who can reach out to the targeted audience. This strategy can be useful

when the barriers to reach the target audience directly is high. For example, foreign actors may use this strategy to reach an audience in another country where foreign influence activities are heavily monitored or banned. A recent Russian operation utilised a fake non-profit independent news outlet called “Peace Data” to engage freelance journalists to write articles on social commentary topics. The use of journalists from a target community or audience is necessary when language barriers make it difficult for foreign actors to penetrate the information environment. Besides being a tool, journalists are a target of influence operations. A report by Trend Micro in 2017 on services offered in the dark web postulated that the rate for silencing or discrediting journalists on social media was around $55,000.

- Media coverage of influence operations has increased significantly after the 2016 U.S. elections. However, the coverage is rarely nuanced, often negative and excludes mentioning that influence operations can also have positive uses. It also belies the complexity of influence operations in a hyper-connected information environment. In this environment, influence operations might originate from a

foreign actor and then amplified by domestic actors. Furthermore, some interventions such as legislation to counter disinformation can backfire.

- Information and communication technology (ICT) is making the media more crucial to influence operations. It increases the speed that news can reach the audiences, hence upping the newsrooms' pressure to break real-time stories. It makes mass communications multi-directional, allowing influence operations actors to obfuscate their origins while engaging the target audience in the form of participatory propaganda. It also makes it easier for foreign actors to engage journalists in other countries and launder content.

- The media is crucial to the health of the overall information environment. Media outlets and journalists are essential nodes in the complex and interconnected system that characterises this environment. Politicians and citizens would be more vulnerable to influence operations if influence operations are successful in ‘infecting’ media outlets and journalists with deleterious ideas. Therefore, governments and the technology sector should provide more help to the media outlets in countering influence operations. Journalists, including students in

journalism schools, should receive more training to keep them informed about influence operations. Governments should provide their citizens with better education on digital literacy to better understand the information environment. Instead of leaving it to social media companies to determine how their platforms are used, the government and the society can help these companies determine what behaviour is unacceptable and should be countered.

The panel sessions of the workshop are captured in the conference report with speakers identified. Q&A discussions are incorporated without attribution.

# About the Centre of Excellence for National Security

The **Centre of Excellence for National Security (CENS)** is a research unit of the S. Rajaratnam School of International Studies (RSIS) at the Nanyang Technological University, Singapore. Established on 1st April 2006, CENS's raison d'être is to raise the intellectual capital invested in strategising national security. To do so, CENS is devoted to rigorous policy-relevant analysis across a range of national security issues. CENS is multinational in composition, comprising both Singaporeans and foreign analysts who are specialists in various aspects of national and homeland security affairs. Besides full-time analysts, CENS further boosts its research capacity and keeps abreast of cutting edge global trends in national security research by maintaining and encouraging a steady stream of Visiting Fellows.

For more information about CENS, please visit www.rsis.edu.sg/research/cens/.

# About the S. Rajaratnam School of International Studies

The S. Rajaratnam School of International Studies (RSIS) is a professional graduate school of international affairs at the Nanyang Technological University, Singapore. RSIS' mission is to develop a community of scholars and policy analysts at the forefront of security studies and international affairs. Its core functions are research, graduate education and networking. It produces cutting-edge research on Asia Pacific Security, Multilateralism and Regionalism, Conflict Studies, Non-Traditional Security, International Political Economy, and Country and Region Studies. RSIS' activities are aimed at assisting policymakers in developing comprehensive approaches to strategic thinking on security and stability issues in the Asia Pacific.

For more information about RSIS, please visit www.rsis.edu.sg.

## About the National Security Coordination Secretariat

The National Security Coordination Secretariat (NSCS) is a unit under the Prime Minister's Office responsible for national security planning and coordination. NSCS works with agencies and stakeholders to develop, co-ordinate and implement Singapore's strategies to address national security concerns. NSCS also works with agencies to anticipate and identify emergent security risks, and to build up capabilities and resources to deal with these. NSCS departments include Geopolitical Security, Resilience Security, Civil Security and National Security Risk Planning.

Please visit http://www.nscs.gov.sg for more information.