CENS Workshop on DRUMS 2023: The Evolving Landscape of DRUMS

Event Report

30-31 October 2023

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

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Event Report

CENS Workshop on DRUMS 2023: The Evolving Landscape of DRUMS

30-31 October 2023
PARKROYAL COLLECTION Marina Bay

TABLE OF CONTENTS

Organised by:
Centre of Excellence for National Security (CENS)
S. Rajaratnam School of International Studies (RSIS)
Nanyang Technological University (NTU), Singapore

Rapporteurs:
Asha Hemrajani, Tan E-Reng, Sean Chua, Eugene Tan, Shantanu Sharma, Jermane Soh, Diana Ho, Yasmine Wong

Editors:
Dymples Leong and Xue Zhang

EXECUTIVE SUMMARY

1. On 30 and 31 October 2023, the Centre of Excellence for National Security (CENS) conducted its annual DRUMS (Distortions, Rumours, Untruths, Misinformation and Smears) workshop with the theme “The Evolving Landscape of DRUMS”. This was the eighth DRUMS workshop that CENS has organised since 2017, including the three closed-door webinars during the COVID-19 pandemic.

2. Fifteen international and local speakers covered topics on the evolution of disinformation and misinformation, safeguarding electoral integrity, strategic communications, sophisticated generative AI tools, counter DRUMS measures, and information, media & health literacy. Over 180 persons from government agencies, academia, and non-government organisations (NGOs) attended the workshop and participated actively in the Questions and Answers (Q&A) and syndicate discussion sessions.

3. Foreign Information Manipulation and Interference (FIMI) and the Information Manipulation and Interference (IMI) from within respective countries featured prominently throughout the panels. The keynote speaker introduced how NATO (North Atlantic Treaty Organisation) challenges disinformation and hostile information. A common taxonomy, such as the concept of FIMI, would be a good potential way to identify and share lessons across organisations.

4. Panel 1 speakers pointed out that strengthening public trust in institutions and safeguarding electoral integrity remain key priorities, as distrust in electoral processes and media can contribute to the spread of propaganda. Panel 2 speakers discussed the ways in which governments can support agencies and civil society in reinforcing and improving resilience, a constructivist approach for strategic communications, as well as developing countermeasures against full scale information warfare. Panel 3 speakers elaborated on the role of generative AI in influencing the adaptation of fact-checking and reshaping journalism both economically and editorially. Collaboration between governance bodies and tech experts are required to safeguard the integrity of information.

5. Panel 4 speakers noted that adopting counter-DRUMS measures as rhetorical strategies to discredit media scrutiny, attack opponents, and challenge dissident perspectives represent a common political trend on a global scale. Panel 5 speakers underscored the importance of information, media, and health literacy, elaborated by Singapore's public education, survey findings from Japan, as well as from a journalist perspective. It was also noted that the emergence of open-source tools for journalists and fact-checkers has the potential to contribute to a more informed and discerning public.

6. The 2023 edition of DRUMS received very positive feedback on every aspect, including design and organisation of panels, selection of speakers and topics, operation, and administrative matters. Future editions might consider inviting more Asian (pro-east) speakers for holding a more balanced view, focusing more on political insights and (proposed) solutions, and including (more) speakers from social media and/or big tech companies.

7. This report summarises key points from the speakers' presentations. Key takeaways from the Q&A and syndicate discussion sessions are included at the end of each panel.

KEYNOTE SPEECH – HOW NATO CHALLENGES HOSTILE INFORMATION AND DISFORMATION

Chris Riley, Head, Strategic Communications Unit, Public Diplomacy Division (PDD), NATO

- Chris Riley spoke on how NATO handles hostile communications. NATO is a bureaucratic organization by nature but is learning to adapt quickly to the current dynamic environment. There have been mistakes made along the way; however, the NATO mindset is adjusting to allow adaptation.

- The starting point for NATO definitions of hostile information and disinformation was Russia's annexation of Crimea and Donbas in 2014. It provided the impetus to consider information as a strategic pillar of hybrid war.

- As an example of Russian disinformation, a woman claimed to have seen a boy crucified by Ukrainian nationalists. It was quickly disproved by journalists as the protagonist had appeared as an actress in other videos. However, the damage had been done as the video was viral. The "Lisa" rape case was another example cited of Russian co-ordinated disinformation conducted via multiple channels. Russian TTPs (Tactics, Techniques & Procedures) have evidently not evolved as the same fabricated story was repeated 18 months later in Lithuania. The story failed to gain traction as the press was forewarned.

- NATO has adopted a two-pronged approach to countering hostile information activities: conducting an information environment assessment (IEA) and proactive external communications, allowing resources and information to be shared across hierarchy. The IEA piece consists of four parts: analytics using technology, reporting products tailored for different audiences, capability development, and IEA community building.

- There are four key data pillars in the fight against disinformation, namely 1) traditional media monitoring and analysis, 2) hostile information activities analysis, 3) audience research and social media monitoring, and 4) analysis to inform improved political and military decision making. However, even with these sources, care is taken not to draw a scientific line of cause and effect as contribution and causality are different.

- The long-term approach for NATO involves proactive communications, capability development, capacity and resilience building, co-operation with international partners in addition to prioritising pre-bunking of disinformation. This includes getting ahead of disinformation narratives around NATO's nuclear deterrence exercise to better shape the discussions around the exercise despite internal reluctance. Short-term responses to hostile information narratives require decisions on when to

intervene. The team has adopted a risk-based approach that considers the issue, spread and the assessed impact.

- A common taxonomy, such as the concept of foreign information, manipulation, and interference (FIMI), would be a good potential way to identify and share lessons across organizations.

Key Points Noted from the Q&A Session

- Engagement, collaboration with NATO and other parties to monitor and analyse the disinformation kill chain and inauthentic behavior should be the biggest motivator for private industry in contributing to society.

- Countries other than Russia are challenging NATO in the information space, for example China's increasing capacity to engage in NATO's information space strategically. An example of hostile narratives in a security context coming out of China was the anti-NATO messaging during the Covid-19 pandemic.

- Proportionate responses towards disinformation are crucial. Responding to disinformation in the short-term requires agility and nimbleness to respond effectively. Long-term responses to disinformation involve consistent monitoring and getting ahead of potential threats.

PANEL 1 – SAFEGUARDING ELECTORAL INTEGRITY

Loyalists or Propagandists? Political Propaganda and Mis/Disinformation on TikTok During Malaysia General and State Elections

Nuurrianti Jalli, Assistant Professor, School of Media and Strategic Communications, Oklahoma State University

- Electoral integrity has become a critical issue in Southeast Asia. The dangerous impact of disinformation in political processes has been seen in countries such as Indonesia and the Philippines. In Indonesia, unchecked disinformation has led to significant post-election unrest and violence, as seen in the 2019 protests that resulted in 20 deaths. Disinformation campaigns have been particularly effective due to the use of encrypted messaging apps like WhatsApp, which make it difficult for outside observation and containment. In the Philippines, the elections of Duterte and Bongbong Marcos provide stark examples of how disinformation can shape public narratives and historical perceptions.

- Nuurianti Jalli's analysis extends to Malaysia, where she identified government agencies' involvement in vote manipulation and highlights the manipulation of both voting acts and voters' choices through tactics such as phantom voting, vote miscounting, voting machine tampering, and bribery. Other methods include institutional manipulations through gerrymandering and the misuse of public relations departments for spreading electoral disinformation, as observed in the case of Marcos Bongbong, are also noted.

- Factors such as the lowered voting age in Malaysia and the growing influence of social media platforms like TikTok among young voters have contributed towards the evolving landscape of disinformation. Data analysis of TikTok videos conducted revealed propaganda, disinformation, and hate speech, carrying the potential to significantly influence voter perceptions and behaviours.

- In response to these future challenges such as deepfake technology and AI-driven micro-targeting, proposed strategies include enacting context-sensitive electoral policies aligned with global standards, enforcing accountability for political actors and tech companies, bolstering support for independent fact-checkers, and promoting comprehensive media literacy programs to cultivate a more informed electorate.

Protecting Elections against Foreign Malign Influence: Successes and Failures from Europe

Dominika Hajdu, Policy Director, Centre for Democracy & Resilience,GLOBSEC

- A thorough examination of electoral vulnerabilities was presented, identifying a range of threats to the integrity of democratic processes. Public attitudes, the readiness of state mechanisms, and the conduct of political actors form a triad of areas susceptible to undermining electoral resilience. Key points of concern included the polarization of the electorate, resulting in a divided society with reduced capacity to achieve a democratic consensus, and the eroding trust in institutions that ensure the legitimacy of electoral outcomes.

- The electorate's lack of critical thinking, as a significant weakness, could be exploited through misinformation. The absence of transparency within electoral mechanisms and the insufficient capabilities of electoral bodies to swiftly detect and counteract irregularities further exacerbate these vulnerabilities.

- The presentation also highlighted the shortcomings in strategic communication among state and public institutions, which hinders effective dissemination of information and public engagement during elections.

- The pernicious effects of ignoring external threats to electoral integrity can lead to real-world implications, and government acknowledgment and proactive measures are essential for public awareness and defence against such incursions. To illustrate the real-world implications of these vulnerabilities, Hajdu referenced an incident where the Russian ambassador in Bulgaria expressed a voting preference just before an election—a potential infraction of the Vienna Convention's diplomatic norms; and exemplifies the subtleties of electoral interference and the necessity for constant vigilance.

Key Issues Noted from the Syndicate Discussion

Issue: Bolstering media information literacy is essential to empower citizens to discern and evaluate veracity of information

To enhance the safety of elections, it is crucial to foster societal capacity, ensuring that the electorate is well-informed about the electoral process and the importance of their vote. A critical examination of the election commission is recommended to ensure its integrity and trustworthiness. Drawing inspiration from Taiwan's effective public communication during the Covid-19 crisis, employing humour and engaging methods, could be beneficial in raising awareness about election security threats.

Issue: Oversight by state actors and political monitoring agencies can combat disinformation without exacerbating societal polarization

Addressing the challenge of disinformation in Asia involves recognizing the diverse non-state actors involved, such as political entities, opportunists, and loyalists, who fuel the spread of false information. A collaborative approach with social media companies is necessary to regulate the dissemination of disinformation. For instance, TikTok's proactive stance on removing harmful content exemplifies the potential role of social media platforms in this effort.

Issue: Building trust in public institutions by governments

Classifying and clearly communicating the various threats to national security that arise from polarization can also be instrumental. A strategic move to demonetize disinformation aims to undercut the economic incentives behind its spread. This could involve implementing policies to stop platforms from recommending polarising content, thereby disrupting the business models that profit from misinformation.

Issue: Regulating instant messaging apps such as Telegram in the context of Chinese ideology requires careful consideration of racial and social divisions

While Telegram may not be as widely used as WhatsApp in Malaysia, its influence still necessitates a thoughtful regulatory approach to prevent it from becoming a tool that deepens societal fractures. The objective is to maintain an open yet respectful exchange of ideas without allowing the platform to become a breeding ground for further polarisation.

PANEL 2 – THE GLOBAL STRUGGLE OF NARRATIVES: INSIGHTS FOR STRATEGIC COMMUNICATIONS

The Global Struggle of Narratives: Telling China's Story Well

Una Bērzina-Čerenkova, Head, Riga Stradins University China Studies Centre; Head, Asia Research Programme, Latvian Institute of International Affairs

- A constructivist approach is needed for strategic communications with a renewed focus on psychological concepts such as identity. This is even more critical as states undertake more work under the Foreign Information Manipulation and Interference (FIMI) and the Information Manipulation and Interference from within the country (IMI).

- The FIMI framework offers a solid and holistic methodology which puts the behaviour of the actor at the forefront rather than the content perpetuated. The flexibility of this framework allows for it to be used for hostile behaviour. At the first level, FIMI determines if there is malignant intent and if there is an attempt to manipulate. Other determinants include the presence of coordination and if the attempt is intentional. The FIMI framework also includes civil society outreach.

- FIMI is however an easier problem than IMI. This is especially so when the malicious actors are closer geographically as it makes it harder for states to push back against certain narratives. Added complexity such as when identity of an individual is heavily tied to the narrative of community and such an identity is hard to shake off due to the shame to the individual should he/she disavows the message.

- China has been implicated in several incidents in Europe based off the FIMI framework. China is seen as an aligned communicator with Russia, but that does not mean that Chinese narratives fully mimic Russian ones. Chinese public facing narratives are very close to Russian ones, but strategic communications behind the scenes have often been very different. For now, China's hybrid threat to Europe has not been very visible.

- There are several lessons to be learnt in strategic communications from resilient societies:

- Small countries must avoid issue pairing. Communications should be done based on the issue at hand rather than be dragged into parallel issues.
- De-emphasising issues. Strategic communications have moved from highlighting issues to issues being toned down to avoid negative effects.
- Owning the narrative. States need to allow self-criticism into public spaces as allegations typically come from domestic audiences. This ownership includes owning up to failed policies. Failure to do so creates space for hostile information campaigns.
- Strategic communications can be apologetic in nature. Communications do not have to be a battle, but an explanation on why things are being done in a certain way. There is a possibility of overstepping, but this can be remedied in a democracy.
- People in positions to make an impact on society should not be attacked. Some groups may feel pressured by external forces. These people have an opinion as well, and criticism should be measured.

Countering Russia's Destructive Narratives: The Ukrainian Experience of Information Defence

Anayit Khoperiya, Head of the Department for Countering Information Threats to National Security, Center for Countering Disinformation, National Security and Defence Council of Ukraine (NSDC)

- Due to operational constraints, information dissemination needs to be conducted in a way that prioritises speed. This is especially important to disseminate critical information to the public as fast as possible and not cause disorder in a time of crisis. This means analytical reports are not feasible as time needed to declassify and approve reports would be too long and may not be digestible by the public.
- The level of media literacy and critical thinking is also being raised at the same time. The public is being educated on what is disinformation and how living in an information bubble looks like. Education is being carried out by creating a manual on countering disinformation, conducting training courses for civil servants and the creation of cartoons for children.
    - There are various tools used by Russians to perpetuate disinformation against Ukrainians: Fake letters. Messages offering monetary rewards were received to induce the population to cooperate with the Russians.

- Fake websites. Clone websites from the United States have been found to peddle misinformation in Ukraine. Investigations show that there were few articles contained on the website and the information in the article was made up.
- Fake advertisements. Videos showing fake messages in images were found on telegram channels and X accounts trying to create a false narrative that the Zelensky government is not being supported around the world.
- Fake magazine covers. Images from international magazine covers are being manipulated to present false narratives that do not exist in the real world.
- Cartoons are used to reduce support for the Ukrainian government both at home and abroad (e.g., spread as a French website).
- Telegram Channels. Channels are being created to spread propaganda against the Ukrainian government. This is being done in a coordinated manner with frequent posting and resharing of inauthentic content.
- Bots. False narratives are being spun to paint the Ukrainian as being inhuman, and that atrocities are being committed when they have not.
- Kremlin Propagandists often create narratives for the Russian government, including providing information alibis for the government, interviewing foreign pseudo-experts, and changing the essence of foreign media reporting.

Swedish Experiences and Perspectives regarding Psychological Defence and Warfare

Mikael Tofvesson, Head of Operations, Swedish Psychological Defence Agency

- Swedish psychological defence operations waned towards the end of the 1990s after the collapse of the Soviet Union as it believed that peace would come to Europe. This has now been proven otherwise and the programme had to be revitalised to deal with current threats in an agile manner.
- Cooperation among the different groups in society becomes more important especially as the messaging and sometimes physical activity seeks to divide opinions in society both home and abroad. An example would be the burning of the Koran by a far-right politician in Sweden, an act that most Swedes are not supportive of but have far-reaching impacts on society.
- Russian information operations against Sweden intensified after the Russian invasion of Crimea in 2014. Russian information operations mainly targeted Western responses to the initial invasion, including creating social media

accounts and targeting divisive cultural aspects of society. Content can be created to target any part of society and can be meshed with acts of criminal activity to inflame opinion, especially to influence what the Russian population generally think of the west.

- The modus operandi of Russian information operations therefore needs to be better understood and communicated to reproduce awareness in the general population. Russian information narratives can be non-linear in nature and will make use of all, including far-left and far-right narratives, to create chaos in society.
- There is also a need to challenge narratives that do not fit into the general value system of society, albeit in a less robust and conflict driven manner. Most of these challenges come from states that are generally supportive of Russia. For example, China is seen as a potential threat to Swedish sovereignty through its coercive behaviour, and Iran is seen as a threat to societal safety.

Key Issues Noted from the Syndicate Discussion

Issue: The impact and effects of how people react to malign narratives needs to be understood.

There is no one credible solution to measuring the impacts of these narratives, but it is important to know where the information comes from and the potential for existing cleavages in society becoming wider. Malign foreign disinformation can target issues such as Islamophobia or LGBTQ+ to cause mayhem in society. Such problems are often exacerbated by a population's distrust of their own governments on these issues.

Issue: Hostile actors typically spread narratives that often play on emotional connections to generate maximum effect.

Countering these emotive narratives can be difficult but can be addressed by states toning down the surrounding emotional noise. This can enable people to judge how ridiculous some of the claims really are. States need some level of creativity to determine what will happen next and how their own narratives can be tailored to fit their own cultural or patriotic context.

Issue: Hiding from "whataboutism" can be counterproductive.

The use of "whataboutism" was prevalently employed by the Soviet Union to craft their narratives. Governments cannot tackle this issue alone and requires societal effort. The discourse peddled by journalists, politicians, and influencers need to be transparent, trustworthy, and truthful to counter this narrative. This is best represented in the difference in Russian narratives from the war in Ukraine and the Israel-Hamas conflict.

The narrative of Russia as a peaceful country has lost credibility, and Russia has attempted to use the Israel-Hamas conflict to refocus information attention away from the Ukrainian war. This may also involve other like-minded countries that dislike the United States like Iran, who are known Hamas sponsors.

Issue: Offensive and defensive strategies in narrative formation needs to evolve in line with threats.

States need to think strategically and long-term and identify vulnerabilities in society. This may involve improving media literacy and critical thinking in societies. States should also think about existential issues such as the will to defend a country and its foundational basis, such as national values and system of government. For example, countries such as Russia incentivise sacrifice for the country and have suffered large number of lives lost. Denial of a national existence by another country is also a powerful driver for the development of a national identity.

PANEL 3 – SOPHISTICATED GENERATIVE AI TOOLS: NEW BATTLEGROUND OF DISINFORMATION, FACT-CHECKING AND GOVERNANCE

The ABCs of Future AI-Powered Information Operations

John Kelly, Founder & CEO, Graphika

- The production of high-quality information operations has become cheaper, faster, and easier. As an example, an information campaign of a fictitious attack on Zaporizhzhia nuclear power plant with commonly available generative AI tools was created in under 20 minutes and at a cost of less than USD$4.00. The operation used photorealistic images from different angles generated by DALL-E using simple text prompts, posts on X and Telegram in different languages and fake news articles with eyewitness testimonies created using ChatGPT, and avatars and profile pictures of accounts produced using online image generation websites such as https://thispersondoesnotexist.com/.

- Camille François of Graphika established the "ABC framework" for describing and analysing information operations using three criteria: Actors: Identifying the command and control – Who's doing it?

    - Behaviour: Identifying the tactics and weapon systems being used - How are they doing it (use of bot farms, trolls etc.)?

    - Content: What is the ammunition being used – What is the payload (deepfakes, images, readfakes)?

- With even lower barriers on entry due to proliferation of generative AI, information operations are now being conducted by a much wider set of actors. A social media operation was launched with AI avatars pretending to be Americans speaking in American accented English, expressing support for Burkina Faso's new pro-Russian, anti-French president. Experts initially believed that the campaign was being operated by the Wagner Group, but the operation was eventually tracked back to individual in New York City operating in a private capacity.

- Although behaviours and tactics of information operations have not changed drastically, the advent of AI has allowed scaling-up of operations. The labour efficiencies afforded by business with generative AI have extended to

information operation, radically transforming the economics of operations. The limitations due to personnel budget and local expertise required to produce sophisticated and convincing content online through troll and bot farms have diminished. For example, a recent Graphika report “Deepfake It Till You Make It” highlights use of Synthesia software by Chinese state-sponsored actors to create human avatars using generative AI, posing as news anchor to spread disinformation for less than USD$50.00.

- The implication of generative AI on content has been mixed. Although the content is imperfect where experts and specialized AI-based detection tools can still identify that content has been synthetically generated, the content is still “good enough”. The quantity of the content that can be produced allows for it to be spread widely and generate the emotional impact for which it was created. Even if the content is proved to be generated, it is too little too late. It is unlikely that AI-enabled detection tools would be able to keep up with the quality and quantity of content created using generative AI in the future.

Generative AI Tools for Fact-Checking, Automated Detection of and Debunking Disinformation

Svitlana Slipchenko, Head of VoxCheck Project, Member of VoxUkraine Management team

- VoxCheck was founded in 2016 for fact checking statements made by politicians in EU and Ukraine. It is now a partner of Forbes Ukraine, verifying articles for the magazine, and in partnership with the Ukrainian state TV broadcaster. The fact-checking information generated by VoxCheck amasses almost a million views/day. It has debunked more than 3,000 fakes ranging from antivax misinformation to content related to false content related to the Russian invasion of Ukraine.

- The most prominent case of generative AI disinformation content was a deepfake of President Zelensky capitulating in March 2022. The content was easily identifiable as fake even to an untrained audience due to its exceptionally low quality.

- Another significant case of disinformation using generative AI spreading in Ukrainian information space was videos of the Polish president and former president of EU parliament talking about stopping aid to Ukraine. The videos were real while with audio generated using AI tools. The information operation was identified as a scam to lure in subscribers for monetization of the channels spreading the disinformation. The audios were of low quality

and created using ElevenLabs App, used for editing TikTok videos. The target audience of the operation was low-income Ukrainians receiving aid from the EU and Poland.

- Other Ukrainian start-ups are using AI tools for combating disinformation, supporting businesses and armed forces:
    - Mantis Analytics: A Ukrainian start-up that develops AI models to help government bodies and NGOs see trends in disinformation faster. The company is also organizing a Kaggle competition in collaboration with VoxCheck for faster and efficient detection for false information.
    - LetsData: It is a non-governmental organization for detecting and countering disinformation which also work with businesses to analyse and react to black PR campaigns.
    - Who Are You app: It is an application for Ukrainian armed forces to help identify citizens through documents. It was created by YouControl, a Ukrainian company which works with financial organizations and banks to help prevent financial crimes.

What are News Organisations Doing with Generative AI?

Charlie Beckett, Professor of Practice, Director of Polis, London School of Economics

- Journalism has partly changed due to social media. Social media has transformed how people find out about and consume news and changed news production in newsrooms. The public have also started producing news themselves, consuming information directly, which has resulted in the disintermediation of journalism.
- AI was initially used in newsrooms to perform repetitive tasks such as helping investigative journalists sort and analyse 10,000+ leaked documents. AI is now used by traditional media to overcome the information deluge to provide information.
- Generative AI can positively help journalism - through news gathering, production and distribution. There are huge efficiencies gains for journalists such as faster transcribing and language translation by adopting AI in their workflow. AI can help free-up the resources for investigative work, human interest stories, and real-world reporting.

- A survey of Journalism and Artificial Intelligence published by Polis, LSE highlighted that most newsrooms have experimented with generative AI, and more than 80% of the respondents would use AI in newsroom for fact-checking, content personalization, text summarization, creating chatbots to gauge public sentiment, and for conducting preliminary interviews.

- Risks and ethical challenges surrounding AI needs to be addressed, including discrimination, biases, facial recognition and profiling, legal issues around copyright and the monopolization of AI by a handful of companies.

- The advent of AI also creates risks for news organizations by diminishing trust in news due to lack of quality control of AI generated content. There is also a fear that generative AI will create competition for news organizations.

- Journalists should adapt to the new environment and help create ethical models tailored to the newsroom with human-in-the-loop.

Key Issues Noted from the Syndicate Discussion

Issue: Risk of AI-based detection tools making journalism obsolete

During the coronavirus pandemic it was harder to detect organic communities depicting conspiratorial ideology along with state-sponsored conspiracies that occupied the information space. It is hard to categorize what “bad information” is and even harder to create a solution to tackle it. Defining what constitutes as “bad information” based on production and consumption is harder as it is influenced by organic and inorganic content, such as conspiracies in QAnon communities towards antivax attitudes. There is a requirement for more fact checkers, journalists, and analysts.

Issue: Potential challenges faced by newsrooms

Journalism is formulaic and structured. With the proliferation of generative-AI, a transformation of structured news to platform-based specific content (such as image, video reel, and text) would become significantly easier. This will allow journalists to focus more on content, investigation, and research. This can result in a shift in journalism and create further bifurcation of the news industry into big organizations and smaller niche organizations. Though there are productivity and efficiency benefits of generative AI, it can also create competition for journalists where some might resort to bad practices such as click baiting and controversial content to raise engagement and readership.

Issue: Race between AI-based detection tool and generative-AI produced content

While AI-enabled detection tools are currently working, it might not be possible to evaluate individual pieces of content in the long-term. AI-based detection tools would not be able to keep up with the quality and quantity of content produced using generative-AI. Adversaries creating information operations will have a first-mover advantage in the information space. Although it might get hard to detect veracity of single pieces of content, it might be possible to identify bot networks. Upgrading AI-based detection using operational theory to analyse flow of swarms of media and narratives through networks over time can help differentiate between human and inorganic networks. It is essential to create support for human analysts and journalists for judgement and analysis based on context and history which AI cannot perform.

PANEL 4 – COUNTER-DRUMS MEASURES

Governing Fake News: Social Media Regulations and Freedom of Expression in the Age of Emergency

Donato Vese, Assistant Professor in Law and Economics, University of Pisa

- It is important to consider how to mitigate the negative effects of fake news, while at the same time be wary of the risks of stifling free speech over the course of this mitigative process. Consideration should be given to various factors as to why people credit falsehoods, including informational and conformity cascades, group polarisation, and echo chambers.

- Individuals are generally more willing to believe a piece of information if it is personally relayed to them. There is a latent propensity to believe other people by default, known as "truth bias". Individuals are generally predisposed to believe certain rumours or falsehoods even if they cannot be proven or verified, as they generally have the mentality that the mere fact of the rumour existing implies a certain level of truth to the rumours to begin with.

- This has raised a question on whether falsehoods and fake news should be protected in a society that is committed to upholding the principles of free speech. Three primary concerns surrounding this discourse:

    - Fallibility: If policymakers are allowed to punish people who propagate falsehoods, there is a risk that the overall apparatus used to make these judgments might not be reliable or fair. Policymakers could potentially make politically motivated decisions to censor certain opinions and content that is undesirable to them, deeming it "fake news" merely as a pretext. Fake news policies and laws could potentially be abused to censor "undesirable" discourse. i.e., Dr Li Wenliang was disciplined for "spreading misinformation" in China during the early days of the COVID-19 pandemic, when he attempted to call attention to ongoing issues surrounding the pandemic.

    - Living Truths: In a natural democracy, policymakers should not base their policies around a wholly punitive model to punish proponents of fake news but should ratify policies that empower people to debate,

understand, and learn more about falsehoods and fake news. The benefit of this is that people can become more literate and skilful in recognizing errors and fallacies behind false facts. Having greater literacy about these issues would enable people to correct others with deficits in knowledge and understanding regarding certain topics, as well as people with higher propensity for a herd mentality where they simply believe information without much critical evaluation.

- Chilling Truth: People should not be afraid to speak the truth for fear of facing punitive measures or legal threats.

Counter-DRUMS Measures Reconsidered: The Cases of Chinese Social Media Platforms and Beyond

Jun Liu, Associate Professor, Department of Communication, University of Copenhagen

- An alternative, reflective approach towards understanding DRUMS should be considered. In the context of political arenas, DRUMS can be used as a pretext for political accusation and manipulation, where information is censored for politically motivated reasons, rather than for being objectively false. Counter-DRUMS measures could be misused to dismiss political criticism by opponents (e.g., Donald Trump).

- Michel Foucault's 1980 theorizations on truth could be invoked to better parse these issues. Instead of looking at a piece of information and gauging its truthfulness and reliability in silo, one should also consider the context of the society where this piece of information is propagated. Questions must be asked as to what kinds of discourse are generally accepted by a particular society, and how this new piece of information would be received by the society upon its propagation and consumption.

- Fluctuations in public opinion should be closely monitored and charted over time (i.e., by monitoring online spaces, social media etc.). Studies regarding how to determine what constitutes "truth" and what constitutes "falsehoods" within a certain society must account for and contextualise themselves to that society's particular "regime of truth". It is important to always keep in mind who has the authority to define what "truth" means in each society. i.e., different Chinese social media platforms available can serve as examples to illustrate this divergence in authoritative actors involved in defining and refuting rumours.

- For Weibo, the official anti-rumour accounts are run by Chinese government agencies. The government is, in this case, the authoritative body that defines "truth". Weibo's anti-rumour accounts essentially consolidate government discourse on DRUMS. On WeChat, anti-rumour accounts are conversely run by professional communities outside the Chinese government's immediate purview. Fact checking activity and discourse originate mainly from public sources.

- The challenge posed by the dominant role of the government in certain anti-rumour operations within the social media space might compromise the authority and credibility of counter-DRUMS measures, creating situations where counter-DRUMS measures might backfire. In China, government-initiated discussions on DRUMS have been observed to generate the opposite of the intended effect, where the propagation of falsehoods and DRUMS content experienced an increase, instead of a decrease, upon the initiation of counter-DRUMS discourse by the government. An example of this was the Red Yellow Blue scandal, concerning multiple allegations of child abuse in a Beijing kindergarten, where photos of allegedly bullied children went viral, being shared widely by parents. Social media users, as opposed to being discouraged from discussing the rumours upon the initiation of counter-DRUMS activity by the government, were instead encouraged to talk about these rumours even more.

#ForFreedom – Project Combatting Disinformation during Russian Aggression against Ukraine

Giedrius Sakalauskas, Director, Res Publica - Civic Resilience Centre

- #ForFreedom is a project for combating disinformation, promoting truthful information over the course of the ongoing Russian aggression against Ukraine. Its main target is the Russian-speaking population, across various countries and territories, including Russia itself.

- Traditional models and approaches to countering DRUMS involve debunking—fact-checking and correction efforts. However, these approaches do not always address the root of the existing problems, and often come too late, as they can only be deployed in a reactive capacity (i.e., in response to an influence operation, the detection of a piece of falsehood etc.).

- Debunking continues to be important, but it should be done in a professional way. Otherwise, it could potentially cause more harm instead

of fighting disinformation. Constructive storytelling provides an alternative method in combating DRUMS.

Key Issues Noted from the Syndicate Discussion

Issue: Counter-DRUMS measures could be used as a means of political and social control, in both authoritarian and democratic regimes.

The spreading of DRUMS could conversely be something that can act as a model of contention, resistance, and disobedience to delegitimize the authorities. DRUMS could be weaponized as a means of countering political oppression, particularly in authoritarian regimes. The importance of drawing a line between countering DRUMS and stifling freedom of speech should be emphasised, e.g., the European Union was criticised for introducing an "Orwellian ministry of truth" with its Digital Services Act.

Issue: Mass belief in DRUMS could be attributed to deeper problems with how citizens relate to the political world in their specific socio-political contexts.

The dominant role of governments in certain anti-rumour operations within the social media space might compromise the authority and credibility of counter-DRUMS measures, creating situations where counter-DRUMS measures might backfire. i.e., in China, government-initiated discussions on DRUMS have been observed to generate the opposite of the intended effect, where the propagation of falsehoods and DRUMS content experienced an increase, instead of a decrease, upon the initiation of counter-DRUMS discourse by the government.

Issue: How to measure the effectiveness of counter-DRUMS campaigns

There are both quantitative and qualitative measures of effectiveness when it comes to these campaigns.

- Quantitative measures could include the measurement of numerical, quantifiable data taken from social media platforms. For example, the counting of the number of campaign hashtags being used, as well as the total number of reactions on social media posts on platforms such as Facebook.

- Qualitative measures could include seeing whether real-world actions have been taken as a clear result of certain campaigns. For example, campaign for Walmart to stop selling products with Soviet imagery by Lithuanian activists concluded with Walmart pulling the offending products

from their shelves, marking the success of their campaign.

PANEL 5 – INFORMATION, MEDIA, AND LITERACY FOR COMBATING DRUMS

Nurturing an Informed Citizenry in the Age of Generative AI

Damian Wang, Senior Librarian (Outreach), Programmes and Services, Archives & Libraries Group, National Library Board

- Launched in 2013, the S.U.R.E. (Source, Understand, Research, Evaluate) is the National Library Board's (NLB) flagship information literacy programme. The programme works with partners to promote digital literacy, tailoring content to fit the demographic needs of the audience and strives to make digital literacy appealing to the masses.

- For instance, the S.U.R.E. campaign for adults feature collaboration with SGSecure and the Ministry of Culture, Community and Youth (MCCY); while the S.U.R.E. courseware for younger audiences feature videos, infographics, and parent toolkits, carried out through partnerships with schools. For seniors, the focus is on tailoring the content by language and dialects to engage them more meaningfully. NLB also collaborates with SG Digital Community Hubs' digital ambassadors to engage the elderly with content through digital clinic sessions.

- The S.U.R.E. program is focused on pre-bunking, which involves being educated about misinformation before encountering it. It relies on instilling a critical and awareness-driven mindset to build media literacy. However, the subject matter has evolved and so has the content, with a focus on education on generative AI. When S.U.R.E. started, it dealt mostly with hoaxes; however, after the 2016 US (United States) Presidential Elections, there was a rise of "fake news", exacerbated by the COVID-19 pandemic of 2020. The nature of fake news evolved in 2022 with the Ukraine War, and more recently with generative AI and the Israel-Hamas Conflict.

- With the rise of generative AI content, there is a proliferation of courses on using generative AI programmes, but less focus on the sensitisation of the public to the "dark side" of generative AI. The campaign shows audiences how easily weak guardrails can be exploited by bad actors and how generated images are becoming increasingly convincing. However, imparting the public with knowledge on how to navigate and identify generative AI content remains a challenge.

- The S.U.R.E. campaign features steps to practise digital literacy, however, the public still found it hard to contextualise theory to action. There was a persistent reliance on “common sense” or “gut feel”, and some did not personally feel digital literacy was applicable to their lives. Members of the public were also less able to see the implications of fake news vis-à-vis scams.

Health Literacy in the Fight Against the COVID-19 Infodemic: The Case of Japan

John William Cheng, Associate Professor, Department of English, College of Liberal Arts, Tsuda University

- The ways of countering the infodemic include censorship, fact-checking, and health literacy. The presentation was based on a study conducted in Japan on health literacy and the infodemic. The study found that health literacy reduces belief in COVID-related misinformation, but not COVID-related conspiracy theories. It was also found that individuals who were disproportionately affected by the pandemic, and those who relied heavily on social media for news and avoided mass media news demonstrated lower health literacy and high susceptibility to both misinformation and conspiracy theories.

- Japan was an outlier when it came to COVID: comparably, the country did not have very stringent restrictions, but it also had one of the lowest death rates. However, Japan was not immune to the infodemic. The study showed that 72% of interviewees came across COVID-related fake news, 15% tried unproven preventative measures, and 15% bought extra supplies. Existing literature also showed that fewer people in Japan were affected by fake news despite being lower literacy rates. Groups like JAnon, the Japanese branch of the American far-right group QAnon, were also responsible for spreading vaccine disinformation.

- The study revealed that compared to those who were “health literate”, “sceptics” (higher susceptibility to conspiracy theories) showed higher levels of conspiracy mentality, racist attitudes, low governmental trust, and high social media news consumption. On the other hand, compared to those who were “health literate”, the “misinformed” (higher susceptibility to misinformation) demonstrated all the above factors, combined with low mass media news consumption and were more affected by the pandemic.

- It was also found that health literacy was helpful in countering the infodemic

as it reduced belief in misinformation, but it had no direct effect on conspiracy belief. The factors contributing to this were multi-fold, and a few psychological reasons were posited. Misinformation can be corrected easily because it targets System 2 Thinking, predicated on “slow, conscious, rational” thinking. Conspiracy theories on the other hand, are predicated on System 1 Thinking, which is “fast, intuitive, emotional”. Conspiracy theories capitalise on fear, anxiety, xenophobia, and evoke a general sense of amusement that misinformation might not.

- While health literacy is important, practitioners must think of the psychological barriers and the accessibility of relevant programmes. For example, in Japan, health information has been adapted into cartoons in efforts to debunk fake news and make information more palatable to the public.

Media Literacy and Reporting on Disinformation

Jane Lytvynenko, Freelance Reporter

- The presentation discussed the role of journalism in media literacy. Misinformation is often examined as information itself, contextualisation to the ecosystem within which it exists. The media manipulation cycle provides a nuanced framework for the analysis of misinformation.

- The first phase of the media manipulation cycle involves campaign planning, while the second utilises social media to campaign dissemination. The third stage is critical, where the campaign enters public consciousness and public discussion. At this stage, the information becomes even more relevant when politicians, influencers, journalists etc. amplify said false information. The next stage involves mitigating responses by social media companies, state authorities, or fact checkers to mediate the false information. The last stage is where mis/disinformers adjust to mitigation measures that have been introduced.

- Examining disinformation as abstracted from the network also exacerbates the tendency to speak of disinformation as nebulous. However, disinformation is created by people, and ultimately it is a human problem. Understanding disinformation as a network allows acknowledgement of the fact that the disinformation industry has adapted with the professionalisation of disinformation. There is an increasing wealth of evidence of private firms managing manipulation campaigns, with state actors spending on contracts with such firms for computational propaganda services. States have also

started using official channels to spread false information.

- Bad actors are also adapting, with the environment in 2023 being very different from that in 2020. Part of the reason is that social media networks are changing their policies on how they are policing false information. For example, Meta, when it was known as Facebook, used to have Crowd Tangle, which allowed researchers, academics, and journalists to determine how information is flowing across the platform. However, the tool was used during the US Capitol riots revealing how bad actors were using the platform to plan and build an audience. Following an expose by journalists, Facebook discontinued access to the service.

- Media literacy must include a whole of society grassroots approach. Research has shown that community contexts are crucial in how information is understood. News deserts, for example, are places where there is less traditional media coverage and as such, social media fills that void of the need for local news. As community news faces crises globally, lawmakers need to consider the protection of community news as they are better embedded in audiences' information environment and have more nuanced understandings of their demographics. Fact checking is also facing a crisis and is declining, and registration of new fact-checking organisations is falling. Meta funds about half of fact checkers worldwide, and this reliance on private investment means that if companies decide to withdraw investments, professional fact checking services will continue to decline unless governments step in with fundings or grants.

  Finally, journalists have the ability to educate the public on the cycle and continuity of media manipulation and disinformation that academics might not have access to. They can explain to the public the actors and methods in the disinformation cycle. It is essential to raise public awareness on how information has evolved.

Key Issues Noted from the Syndicate Discussion

<u>Issue: Participation in fake news as entertainment</u>

- Looking at hoaxes as a form of fake news, one of the most widespread hoaxes observed are death hoaxes. For example, there was a hoax spreading that Daniel Radcliffe had died, originating from teenagers in a chatgroup who just did it "for fun". The concept of fun derived from deception is not often talked about when examining hoaxes. With the younger generation, although many can discern real from fake information, they still

share it for entertainment.

- Conspiracy theories have also become part of popular culture, and in fact, conspiracy thinking is necessary: we need to have a healthy amount of scepticism, but it becomes problematic when conspiracies are transposed onto reality. As such, the goal is not necessarily to stop individuals from sharing conspiracy theories, but to understand the difference between reality and theory. However, this remains challenging as conspiracy theorists are not seeking out education and are extremely sceptical of the establishment, as such, it may be difficult for government or official programmes to resonate withs them.

Issue: Comparing private sector and state initiatives to counter disinformation

- In Singapore, announcements of new legislation raised concerns from the public over the potential coercive nature of the legislation (e.g., POFMA). Entities like libraries have a reservoir of goodwill. As such, they can introduce a soft touch to balance out legislation, and both are important.

- The Canadian example serves as a case study on how legislation targeting social media platforms were counterproductive. Recently, to rejuvenate the traditional media industry, the Canadian government mandated that social media companies had to pay news providers (of platform, which was contrary to the intent of the legislation. Regulators do not necessarily have to target the information itself, since there is an entire ecosystem that supports the information disorder. For example, social media advertisements are not held up to the same standards as traditional advertisements, and the lack of algorithmic transparency exacerbates the problem. This is because targeting the speech itself may have a chilling effect on journalism.

- Another issue is the public perception of state initiatives. Research has found that in some countries, the more people perceive private factchecking abilities to be reliable, the less supportive they will be of regulation by authorities.

- Despite the importance of private factchecking, there is a shortage of resources as investments and grants have slowed, and surviving by online advertising revenue alone is unrealistic. Besides waning interest, there is a very high burn out rate for fact checkers as it is a highly demanding job, especially in breaking news situations and in the face of online harassment. As such, it is important for governments to see the value in private fact checking services and allocate resources accordingly.

Issue: Political affiliations and beliefs

- There is strong observation of beliefs mapping onto partisan lines. From a newsroom point of view, in the Western context, there are political believers who remain unconvinced by fact. These individuals are very rarely a majority and their votes do not always find support in the political system.

- The privilege of a newsroom, however, is covering a wide variety of topics, many of which are not overtly political. Furthermore, the most basic journalistic responsibility is informing audiences with well-represented information based on a diversity of sources. As such, journalists can speak to a diverse audience of people on issues beyond politics.

- Community journalism is important as day-to-day activities can unite communities; for example, open-source reporting using satellite information and videos posted online provide a beat-by-beat breakdown of story development. Such methods makes it easier to tackle as evidence is presented without overtly making a political point. Although such means of reporting has also proven to fall into the trap of unverified and conflicting information, this can be corrected by good practices. For example, as more information emerges, news presented should adapt and correct itself, and explain to its audience what gave rise to the mistakes made during reporting. As such, a lot of tools can be utilised to unite people across the political spectrum.

- In East Asia, there are historical factors that give rise to “touchy” issues surrounding Japanese, South Korean, and Chinese relations. The threat facing these issues now is that conspiracies and misinformation are passing down hate from one generation to the next. Because these issues are historical and highly divisive, debunking does not seem as effective.

Issue: News media and the amplification of fake news

- Original fake news is often obscured within a niche community. But fact-checking surfaces it to mainstream public consciousness. When journalists cover something, they are amplifying certain narratives, and strategic silence can be used to dampen certain narratives. For example, the strategy of strategic silence was used in the 2016 elections, albeit unsuccessfully. As such, instead of targeting the narratives themselves, alternative solutions include publishing a list of dis-informers known to peddle misinformation.

About the Centre of Excellence for National Security (CENS)

The **S. Rajaratnam School of International Studies (RSIS)** is a global think tank and professional graduate school of international affairs at the Nanyang Technological University, Singapore. An autonomous school, RSIS' mission is to be a leading research and graduate teaching institution in strategic and international affairs in the Asia Pacific. With the core functions of research, graduate education, and networking, it produces research on Asia Pacific Security, Multilateralism and Regionalism, Conflict Studies, Non-traditional Security, Cybersecurity, Maritime Security and Terrorism Studies.

CENS is a research unit of RSIS at the Nanyang Technological University, Singapore. Established on 1 April 2006, CENS *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.

For more details, please visit www.rsis.edu.sg and www.rsis.edu.sg/cens. Join us at our social media channels at www.rsis.edu.sg/rsis-social-media-channels or scan the QR code.

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

Nanyang Technological University, Singapore
Block S4, Level B3, 50 Nanyang Avenue, Singapore 639798
Tel: +65 6790 6982 | Fax: +65 6794 0617 | www.rsis.edu.sg