SECURITY IN A POST TRUTH WORLD

Event Report

2-3 November 2022

Centre of Excellence for National Security

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

NSCS NATIONAL SECURITY COORDINATION SECRETARIAT

Event report: 2-3 November 2022

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

Rapporteurs: Antara Chakraborthy, Asha Hemrajani, Dymples Leong, Eugene Tan, Yasmine Wong, Zhang Xue

Editors: Sean Tan and Shantanu Sharma

Contents

Executive Summary

On 2 and 3 November 2022, the Centre of Excellence for National Security (CENS) organised an in-person Distortions, Rumours, Untruths, Misinformation and Smears (DRUMS) workshop titled "Security in a Post Truth World". As the first in-person CENS workshop on DRUMS since 2019, the 2022 edition reflected the significant changes in this field over the past three years, addressing the challenges posed by a world where disinformation is endemic and exploring new methods to mitigate these challenges. This was the seventh annual DRUMS workshop that CENS has organised since 2017, including the three closed-door webinars conducted online during the time of COVID-19 pandemic restrictions.

18 speakers from institutions in Australia, Indonesia, Singapore, Thailand, the United Kingdom, and the United States shared their insights over six panels and one special technical presentation. Speakers (and panels) examined disinformation and misinformation, online harms, hate speech, influence operations, fact-checking and regulation. Workshop participants included members of the Singapore civil service, the private sector, and academia, who participated actively in the syndicate discussions and Q&A sessions with the speakers.

Of particular note was the focus on online harassment in the context of gender and misogyny. The sharp rise of Online Gender-Based Violence (OGBV) during the pandemic prompted discussion with regard to laws and regulation designed to protect OGBV victims. More broadly, the workshop gave participants the opportunity to examine the regulatory approaches toward online platforms in different countries. However, at the same time, speakers were also keen to identify misinformation as a moving target, and emphasise the importance of providing users with the necessary tools to empower their own online decision-making.

The workshop's "strategic communication" panel saw particularly fruitful discussion, reflected in positive feedback from participants. Examining strategic communication approaches to the problem (including AI, edutainment and social media literacy) will remain important in future workshops, particularly as the information ecosystem continues to evolve. The workshop's special technical presentation, which demonstrated how algorithms are used to manipulate social media messages and communities, highlighted the saliency of this issue.

Based on the overall positive feedback received from the 2022 edition, the above issues should continue to inform CENS' DRUMS workshops. Future editions must also consider the ongoing- and in some cases, unprecedented- advances in Internet communications technology and machine learning, in order for CENS to remain at the cutting edge of these issues. Besides some of the issues listed above, also deserving focus at future events are open-source intelligence and social network analysis tools, which can assist our research and aid policymakers in Singapore to better scale their responses to influence operations and information threats across the board- and in turn, successfully mitigate the impact of malicious online activity.

This report summarises key points from the panel speakers' presentations. Key takeaways made by participants during the syndicate discussions and Q&A sessions are included at the end of each panel section.

Welcome Remarks

Shashi Jayakumar, *Head, Centre of Excellence for National Security (CENS), RSIS, NTU*

- Dr Shashi Jayakumar, Head of the Centre of Excellence for National Security (CENS), opened the workshop by thanking the speakers, distinguished guests and participants. This year's DRUMS workshop marks one of CENS' first in-person events in nearly three years.
- In the past three years, the issues CENS studies in the influence operations, foreign interference and disinformation spaces have morphed; what seemed to have been a simple set of issues just three years ago have now evolved into something much more multifarious, complex and interlinked.
- It is now essential to study subversion, OSINT, and social media and technological developments in these spaces (to name a few topics), as opposed to simply reading, writing and thinking within our traditional silos.
- Furthermore, events during the pandemic have brought a tangible sense of acceleration to what is already a very combustible mix of issues. It is therefore ideal to have perspectives from both the East and West at this workshop, along with a presentation on gender. The tensions between different approaches can themselves be illuminating, as they provide an indication of where different countries are at when it comes to making tough calls.

Panel 1- Content Moderation on Encrypted Platforms

Kiran Garimella, *Assistant Professor, Rutgers University*

- This presentation focused on the presence, weaponization and combating of misinformation on WhatsApp, an end-to-end encrypted closed messaging platform.
- Unlike on open platforms such as Twitter and Facebook, it is impossible to carry out top-down content moderation on WhatsApp, as end-to-end encryption prevents messages between users from being viewed by a moderator. Bottom-up and offline approaches to combating misinformation on WhatsApp are therefore required.
- WhatsApp is a platform with design elements to handle misinformation. The decentralised nature of the platform, as well as the absence of a broadcast function and algorithms to amplify and personalise messages, makes it difficult for malicious actors to weaponize misinformation on a large scale. WhatsApp's peer-to-peer design also ensures user trust, as they receive information from other users they already know.
- WhatsApp groups feature heavily in Indian political parties' strategies. These groups are loosely organised around political movements and policies, allowing Indian parties to run hundreds of thousands of groups. Disinformation originating in these groups is then used to smear opponents. These smears may take on the form of viral memes, cheap fakes, and photoshopped images.
- While hate speech on WhatsApp is typically not explicit, its impact on minority communities in India is far greater than the volume suggests. Most hate speech is generated from the fear and othering of these minority communities. Incidents and narratives that have been used to stoke hate include India losing a cricket match to Pakistan, and speculation about Islam replacing Hinduism as the majority religion in India.

- WhatsApp users are also mobilised and actively engaged in cross-platform posting. For example, hashtags on Twitter are hijacked by WhatsApp users, who post ready-made tweets shared by operatives from political parties. Such campaigns to manipulate trending Twitter hashtags occur daily. By undertaking these campaigns, political operatives hope to game algorithms into further amplifying these hashtags- and hence reach more users.
- Crowdsourcing for tipoffs about misinformation in these groups is a useful way to build a bottom-up approach to content moderation on WhatsApp. These tipoffs are then quickly fact-checked for the informants. Technology is used to scale up these fact-checking efforts, which are also now available in six different languages.

Information Operations on Wikipedia

Abhas Tripathi, *Disinformation Manager, Wikimedia Foundation*

- There have been many different attempts to conduct information operations on Wikipedia, not unlike those on openly-hosted social media platforms.
- State-sponsored operatives seek to change the narrative and tone of Wikipedia articles. This process, however, is not easy, and would require a well-motivated, well-funded, and long-term campaign.
- Wikipedia is an exercise in verifiability, as opposed to an exercise to capture the truth. It relies on unpaid volunteer editors to post, edit, and verify content on entries as a collaborative effort. The process of information verification also allows a core group of editors to determine the suitability of new editors wishing to enter the group. The bar for new editors is set high in order to maintain standards, as well as a level of commitment. While editorial groups retain significant influence over the direction and editorial standards of their community projects, Wikipedia itself does not maintain specific guidelines over how entries and content are moderated.
- Malicious actors conduct information operations and contribute false or misleading information through a variety of means:
    - **Source-related operations.** Malicious actors contribute content and articles that do not conform to community guidelines of verifiability and reliability. This includes inserting claims without sources, fictional sources, biased sources, and relying on circular reporting.
    - **Sock-puppeting.** Editors' identities are manipulated, and multiple accounts are created to give the impression that a large group of editors are behind changes to a particular article.
    - **Astroturfing.** Groups purportedly claiming to be grassroot organisations seek the removal of certain articles, when they themselves are in fact funded by private or political interests.
    - **Undisclosed paid editing and conflicts of interest.** This breaches general Wikipedia guidelines stating that individual editors are required to remain neutral or declare interests when editing a piece.
    - **Project Capture.** A group of users assume complete control over projects- including the community process- and the content that is to be uploaded onto Wikipedia. Editors that are unaffiliated with the group or hold different viewpoints are often harassed and abused.
- State-based actors have previously threatened to label Wikipedia as a fake news source and their editors with jail sentences when verified information uploaded by community editors did not necessarily align with the narrative(s) of the state or government in question.

SYNDICATE ROOM DISCUSSIONS RELATED TO PANEL 1

ISSUE: Political parties with deep pockets can mobilise teams to spread disinformation

It is easy for operatives of political parties to create multiple chat groups to disseminate messages to their supporters. Teams of staff are hired and assigned to these groups, in order to ensure that these messages are widely broadcast. This practice is especially prevalent in states where cheap labour is readily available. In response, messaging platforms have instituted measures to limit the re-sharing of these messages – whether legitimate or otherwise – but remain unable to moderate shared content due to end-to-end encryption.

ISSUE: Community-based fact checking in hierarchical societies

While platforms can highlight community-based efforts to improve digital literacy as a solution to combat disinformation, it remains difficult for fact checkers to correct disinformation spread by their own family members- especially if these are elders in the family. Correcting such disinformation may cause internal strife within families and result in ostracism of fact checkers.

ISSUE: Governments all around the world face tension over the concepts of free speech and harm

It is easy to typecast government policy either into the western camp of free speech or the Asian camp of harm prevention, but there are limits to both ends of the spectrum. Governments that are purportedly more inclined to harm prevention are also more likely to ask for more meta-data from social media companies, and for backdoors to be built into platforms. The flip side is that these backdoors may also be used for political purposes. States are becoming increasingly assertive and sensitive over content posted on social media platforms.

ISSUE: States are trying to manipulate the Wikipedia editorial process

States are trying to game the Wikipedia editorial process by inserting their own narratives and choice of editors into Wikipedia communities, and in turn intimidate communities by threatening physical harm to volunteer editors. There have also been instances whereby editors were forced to relocate to a different country for their safety.

ISSUE: Transparency and verifiability are strong tools to combat disinformation but require a strong community for the process to work

Displaying a timeline and the edits made on a Wikipedia post can highlight any editorial bias. This process is openly verifiable, and nuggets of information may be challenged at any point in time by anyone. It is up to the community to accept any changes either as verified information or a spun narrative by a paid interest group. The vetting process to join editorial communities is also long, and these individuals are given privileges on a tiered basis. There is a robust process of vetting and gaining community acceptance, making it difficult for unmotivated and impatient individuals to change the narrative.

Special Presentation: Influence and Coordination: Detecting Interesting Activity in Social Media

Kathleen Carley, *Professor, Carnegie Mellon University*

- This presentation showcased the capabilities of tools developed by Computational Analysis of Social and Organizational Systems (CASOS) lab for the detection, monitoring, and analysis of information operations in social media.
- Influence on social media involves the shaping of narratives and communities. Narratives are shaped by countering existing narratives, developing new narratives, or by emotional messaging. Communities are shaped by creating or removing groups, or by influencing community leaders.
- Early indicators of influence on social media include surges in the creation of new accounts, and changes in trending conversations and activity. However, these early indicators are limited to signaling that influence operations may occur- and are hence unable to discern the exact narratives and communities that are being shaped.
- To mitigate the shortcomings of early indicators and to empirically assess the impact of social media manoeuvres, the BEND framework helps to characterise the intent behind such manoeuvres, the types of manoeuvres involved, the actors involved, their target communities, and their impact on both communities and narratives under 16 distinct categories.
- Eight categories under the BEND framework concern manoeuvres to manipulate narratives in the knowledge network that are being spread, and the techniques used to spread them. Similarly, the remaining eight categories concern manoeuvres to manipulate social communities, their leaders, and connections within these communities.
- Bots, trolls, and cyborgs are used in these manoeuvres in order to create and deepen polarisation, by joining groups, boosting polarising messages, and backing opinion leaders on opposing sides of a narrative. To orchestrate a larger coordinated movement, multiple polarising issues may be coalesced together. As these groups become larger and denser, this also creates echo chambers, leaving users especially vulnerable to these manoeuvres.
- Trolls are often the source of hate speech amplified by bots to bridge communities, which in turn connects new audiences to speakers promoting hate-speech. This tactic was observed in Pennsylvania (US) pro-reopening groups during COVID-19, which were themselves connected to QAnon, anti-vaccine, and Covid conspiracy groups. To create further polarisation and disagreement, the pro-reopening groups contained coherent messaging, whereas the anti-reopening groups were targeted with incoherent and uncoordinated messaging.
- The CASOS analytics pipeline involves data collection, followed by computational linguistics and image analytics. The next steps involve hate speech detection, troll and bot identification using network analysis, followed by stance attribution of the communities detected through the BEND framework.
- The National Academy of Sciences (US) has identified a new emerging scientific area called Social Cybersecurity, which involves hacking users' hearts and minds. This field is at the intersection between social science, computer science, psychology, and applied political theory, and also considers the cyber infrastructure and engineering challenges required for societal protection. Many companies are increasingly consulting social cybersecurity practitioners.

Panel 2: Countermeasures

Co-designing a Mobile-based Game to Improve Misinformation Resistance and Vaccine Knowledge in Uganda, Kenya, and Rwanda

John Cook, *Associate Professor, Monash University*

- This presentation was based on the inoculation theory for countering misinformation in a 2017 PLOS study conducted by John Cook, Stephen Lewandowsky and Ulrich Ecker. The study found that the introduction of inoculation messaging before exposure to misinformation effectively reduced the impact of misinformation on participants.
- This study involved participants in a global warming petition project on misinformation. Researchers found that climate misinformation had a polarising effect on participants. It found that the impact of such misinformation was greater among politically conservative participants.
- The study also found that using explaining techniques in misinformation was effective in mitigating its effects. For instance, showing an initial inoculative message before the introduction of climate misinformation helped to neutralise the effects of the misinformation. Inoculation encouraged participants to critically think and evaluate information relating to climate change.
- This inoculation theory can be applied across multiple topics, and was most recently applied to COVID-19 misinformation. It is therefore crucial to educate users on the numerous techniques within the misinformation taxonomy, as they were deployed during the pandemic to fuel misinformation. Common techniques include the use of fake experts, cherry picking of information and conspiracy theories.
- A useful form of public engagement and education to improve misinformation resilience is gamification. An example is the "Cranky Uncle" game developed by John Cook and Goodbeast in 2020. The game was subsequently updated with a COVID-19 edition following the pandemic. The "Cranky Uncle" game helped players to spot misleading tactics on vaccine misinformation. It also helped to explain the main fallacies in vaccine misinformation, via a quiz format where players were able to practice critical thinking (and collect points) as they gained experience and awareness on misinformation.
- The game was also localised to suit various contexts, cultures, and languages across different countries. For instance, the game in East Africa contextualised the local reliance on natural therapies and framed discussion on the topic in a way which respected local customs- and yet reduced misinformation on such therapies.

The Weaponised Fact-Check: False Flags and False Checks

Andrew Moshirnia, *Associate Professor, Monash University*

- The weaponization of fact-checking has become increasingly common. This is exacerbated by concerted efforts by misinformation actors to diminish and dilute fact-checking, which subsequently further amplifies misinformation.
- False equivalence is one of the key tactics used to diminish fact-checking efforts. This invokes fact-checkers' fears of being accused of political bias, especially with regard to politically sensitive cultures and/or topics. Fact-checking efforts are thus diminished by attacks on fact-checkers' confidence and consistency. Political pushback to delay fact-checking efforts is also used to delay the immediacy of fact-checking.

- The creation of false fact-checking groups is similarly used to weaponize fact-checking efforts. An example of a false fact-checking group in action involves the Twitter disinformation campaign in response to the assassination of Jamal Khashoggi. To discredit the active campaigning efforts of Khashoggi's fiancé for justice, false online fact-checking groups were activated to discredit and deny the existence of Khashoggi's fiancé, portraying her as an imposter.
- Another strategy involves the impersonation of pre-existing fact-checking groups, which can then significantly damage the credibility of the actual group still in existence. This strategy also involves IP misuse.
- The final strategy involves the dissemination of deliberately shoddy misinformation intended for discovery by fact-checking groups. Such misinformation is specifically manufactured to be discovered and debunked- for instance, as a decoy cheapfake. This content would mirror actual misinformation to be fact-checked, and is therefore designed to pollute the information ecosystem and cause confusion.

Checking on Fact-Checking: What's Effective and What Can Backfire

Edson C. Tandoc Jr, *Associate Professor, Wee Kim Wee School of Communication and Information, NTU*

- A study by the Wee Kim Wee School of Communication and Information (WKWSCI) revealed 314 active fact-checking projects in 102 countries globally – of which, 85 were active in Asia. The study also revealed that at least 41 of these fact-checking projects were independent, while six were sponsored by various governments across Asia.
- Singapore currently has two fact-checking initiatives, namely *Factually* and *Black Dot Research: FactCheck*. While *Factually* is a government fact-checking website*, Black Dot Research: FactCheck* is an initiative of Black Dot Research, a market research company in Singapore. A survey on the perceptions of Singaporeans towards fact-checking initiatives in Singapore revealed that *Factually* was perceived to provide less counterevidence than *Black Dot Research: FactCheck*.
- The survey also found that participants had more positive reactions to *Factually* with regard to fact-checking on political issues, whereby *Factually* was perceived to have greater credibility than *Black Dot Research: FactCheck*. Researchers noted that the style in which fact-checked information was presented contributed to the perception of greater credibility. Information presented in the style of *Factually* – i.e., crisp and concise presentation – was deemed to be more credible and authoritative.
- Public levels of trust in the government also contributed towards the perceived credibility of fact-checking initiatives. Researchers also found that multi-modal fact-checking efforts were likely to be more successful than text-based information. A combination of text and visual based fact-checks was deemed most effective in reducing the credibility of target misinformation.
- The study also revealed that public usage of fact-checking sites did not increase even during the pandemic, when COVID-19 misinformation was most prevalent. Overall, researchers noted low levels of public awareness toward fact-checking resources.
- Finally, the study's results revealed wide regional variations in the usage of fact-checking resources, with 64% and 62% of participants in Manila and Bangkok actively using fact-checking resources, while only 34.5% of respondents in Singapore claimed to use such resources. A plausible explanation for the low levels of usage is the high levels of confidence and trust among participants in their own ability to identify and distinguish misinformation.

Another possible reason is the low relevance of fact-checked information, or a low threat perception level in which misinformation is deemed to be harmless (in line with respondents' own moral judgements).

SYNDICATE ROOM DISCUSSIONS RELATED TO PANEL 2

ISSUE: Considerations regarding target audience for harm reduction countermeasures

There are different forms of harm reduction, ranging from eradicating beliefs in harmful narratives and/or untruths, to ring-fencing these beliefs- which would include preventing the broadcast of such content. The demographic of the audience should be considered to establish realistic objectives. There are three main categories of audiences- convinced, disengaged, and dismissive. To communicate with an audience already convinced about a topic, the goal is to activate them towards the cause; for disengaged, apathetic members of the public, the goal is to convince them; for dismissive individuals, the goal is to change their minds, which often poses the most challenges. In a scenario with limited resources, it may be more efficient to target audiences who are not dismissive and eradicate specific examples of misinformation. Furthermore, within a polluted information ecosystem, the individual is a target as well as a perpetrator of mis/disinformation. In such a scenario, the goal may instead be to interrupt the virality of the untruth.

ISSUE: Trust in state institutions and fact-checking

Recent studies indicate that some fact-checkers engage in fact-checking due to their dissatisfaction with the state of mainstream media in their respective countries. However, the link between trust in institutions and the state and the uptake of fact-checking services remains unclear. In some contexts, a plural and factitious media landscape results in an increased reliance on fact-checking. However, in cases like South Korea, for example, low levels of trust in mainstream media are accompanied by low usage of factchecking sites. In Singapore, interviews reveal that while individuals express cynicism with mainstream media's state-alignment, there is high trust in mainstream media. These high levels of trust in media, as well as the declining concern over fake news in Singapore, contribute to the low uptake in fact-checking services.

ISSUE: Problems faced by fact-checkers

Fact-checkers function on a variety of principles- from the protection of free speech, to advocating for categorically imperative truth, to utilitarianism. The subjectivity of truth presents challenges to fact-checkers, especially when the truth is fraught with nuances and does not present binary choices. However, for fact-checking to be functional and useful to the general public, categorising information as "partially true" may not be as effective in helping the general public understand the information.

Due the increasing speed of content creation, fact-checkers face problems in keeping up with new content. With advancements in technology, some fact-checking services have been automated, which begs the question of the effectiveness of eliminating the "human-element". The use of automation also raises the issue of trust in AI, which is a complex matter.

ISSUE: Using games as a tool for inoculation against misinformation

Inoculation theory was originally inspired by the experiences of captured US prisoners of war during the Korean War. This theory is centred on the idea that individuals can be pre-emptively prevented from indoctrination through inoculation. Games like 'Cranky Uncle' that serve to inoculate audiences against narratives and fallacies employed in mis/disinformation have proven to be successful, with significant increases in ability to identify misinformation techniques post-game. However, knowledge of cognitive biases does not equate to the ability to apply them effectively in real-life scenarios and is not a panacea. What is instead recommended is a combination of approaches, including fact-based approaches, logic-based approaches, and source-based approaches both online and offline. Partnerships with both state and non-state organisations are also important in ensuring a wider reach.

Panel 3: Regulating Platforms

Regulating Social Media Platforms to Protect Human Rights in Southeast Asia: A Civil Society Perspective

Sutawan Chanprasert, *Executive Director, DigitalReach*

- This discussion focused on regulating social media platforms from a civil society perspective- an especially trending topic given the role of social media platforms in civil society today. Drawing from DigitalReach's work as a member of the Southeast Asian Coalition on Tech Accountability (SEACT), it addressed the impact of technology on human rights in Southeast Asia. Amid Southeast Asian governments myriad attempts to regulate social media, it emphasised that social media regulation can- and should- be achieved while preserving human rights.
- Moreover, the concept of 'social media accountability' is itself inextricable from platforms' obligations to uphold the values of human rights and protect democracy. Two notable examples of failed social media accountability in Southeast Asia include Facebook's role in the Rohingya massacre in Myanmar (which subsequently sparked a commission from Facebook itself), and both 2016 and 2022 elections in the Philippines- the former dubbed by Facebook as 'patient zero' of the modern disinformation era.
- While social media companies have complied with government orders to block and/or remove content in line with regulatory frameworks, regulating social media in Southeast Asia necessitates more than just content removal. Despite this, an observable trend in Southeast Asian regulation involves legislation that is solely focused on removing and suspending content (even though social media regulation is ultimately a far more complicated process).
- The structures of social media platforms themselves can also be problematic, whereby policies are too Western-centric to deal with local issues. For example, Western-centric YouTube policies were not inclusive enough to deal with Philippine-related disinformation during the 2022 election. Furthermore, algorithms and business models often reflect Western biases- for instance, by failing to take into account local and non-English cultural contexts- amplifying disinformation and misinformation. The creation of filter bubbles from algorithms and business models further contributes to the spread of misinformation among the platforms' users bases.

- As such, conversations about social media regulation ought to be centred around structural changes and long-term approaches, as opposed to merely directing companies to remove content one after another. A useful first step would involve more transparency around government oversight, particularly the decision-making process with regard to which content should (and should not) be removed. This step would not only help to prevent excessive censorship, but also help to protect Southeast Asian governments' reputations.
- Ultimately, the characteristics of 'good' regulation are very challenging to define. It is recommended for governments to avoid being overzealous in interfering with online content, and to instead regulate at a system and design level. This would form part of a more sustainable process intended to make companies more accountable in the long-term, instead of reactively responding to censorship orders from the government.
- Finally, it is also important for social media platforms to conduct independent audit assessments (under equal partnerships, in order to preserve the independence of the work), as well as for legislation to be designed in such a way as to create safe and inclusive spaces for all social media users.

Lessons Learned from Content Moderation in Indonesia: Towards Content Governance

Sherly Haristya, *Independent Researcher*

- This discussion presented research conducted in June 2022 in collaboration with Article19 and the European Union on the issues of internet governance and personal data protection, and which highlights the importance of a civil society coalition on content moderation and freedom of expression in response to ongoing challenges in Indonesia.
- Indonesia is home to the world's third largest democracy, the largest economy in Southeast Asia, as well as the world's largest Muslim population. Amid orchestrated efforts to further deepen the already well-established historical roots of segregation in the country, questions remain over the capacity for vocal social media companies in the US and China to understand local experiences in such a diverse country (consisting of 37 provinces, and over 300 ethnic groups and local dialects). This problem is exacerbated by smooth political operators who have the ability to craft and amplify 'grey-area' content that is difficult to regulate, which could then lead to violent real-world implications upon being disseminated by paid political influencers.
- As such, social media regulation in Indonesia should aim to strike a balance between freedom of expression and ensuring safety of individuals and the public. There is also a need to examine the local context of the content in question. Although social media platforms initially insisted that there was no evidence linking problematic (e.g., anti-Chinese) content to physical violence in Indonesia, the 2019 Jakarta riots, which resulted in six protestors killed following clashes with troops over presidential election results, disproved this comprehensively.
- A lack of language knowledge by either platforms' AI content moderations systems or human content moderators has also led to the wrongful takedown of legitimate content (e.g., the takedown of videos and livestreams due to misrecognising information). This results in the violation of user rights without accountability and remedies. While platforms tend to agree to moderate content reported by mass users, there is nevertheless limited dialogue between platforms and Indonesian civil society organisations- this is particularly problematic for minority groups. Furthermore, findings from Australian universities (funded by Meta) also suggest problems with trusted partner initiatives, particularly regarding a lack of transparency over the identity of these trusted partners, and how much influence (disproportionate or not) they have.

- One of Indonesia's largest fact-checking outlets, MAFINDO, is keen to emphasise that they are an organisation that promotes freedom of expression- albeit one that must also promote responsibility. While MAFINDO has appealed for social media platforms to remove hoaxes, these platforms have disagreed, demonstrating a lack of consensus and proportionate influence between platforms and civil society organisations.
- Regarding ongoing efforts by civil society groups and social media platforms to form a coalition on content moderation and freedom of expression, it is first necessary to strengthen the internal capacity of civil society groups (so that they have enough knowledge), as well as a need to build trust among different groups given their diversity, in order for these groups to then engage in continuous and meaningful dialogue with powerful stakeholders.
- Three types of legitimacy are essential to this coalition- input (internal structure), throughput (appropriate decision-making procedures), and output legitimacy (perceived external credibility by other organisations). These forms of legitimacy will allow civil society groups to effectively influence wider stakeholders, and make well-advised decisions moving forward.

Mitigating Digital Disruption: Strategies from the Asia Pacific and a Case Study of Combating Disinformation During Australia's 2022 Election

Andrea Carson, *Professor of Political Communication, La Trobe University*

- With regard to mitigating digital disruption, some of the key takeaways from this study are most relevant- that misinformation is a global problem, and that answers rely on a multi-pronged approach and collaboration. Furthermore, it is important to ensure a healthy balance between responsible speech and mitigating harms that come from disinformation, as getting this balance wrong can result in very serious consequences.
- While misinformation refers to the spread of inaccurate content, whether intentional or not, disinformation specifically refers to the spread of inaccurate or deceptive content through *decisive actions* (Gibbons and Carson, 2022). Although there is no single universally-agreed upon definition of either term, both fall under the umbrella of fake news, a harmful phenomenon made even more widespread by the Internet.
- Some of these harms are economic and electoral in nature, however, they can also be intangible- for example, relating to trust in institutions, and an epistemic crisis in the quality of information in the public sphere. This in turn fuels discriminatory behaviours causing civil unrest (such as the Capitol riot on 6 January), and further polarisation and policy unresponsiveness.
- As such, the above harms pose a threat to national security, and are hence a top-level policy concern for governments- however, governments (who have a duty to protect their citizens) face a challenge to address this without overreach. An estimated 83 countries have used the pandemic to justify violations of free speech and peaceful assembly, occasionally resulting in a chill effect.
- Co-regulatory responses range from the voluntary (e.g., Australia 2021 misinformation code of conduct) to the mandatory (e.g., the EU's Digital Services Act). Partially free countries have tended to adopt legislative approaches and state-sponsored initiatives.
- While EU initiatives were initially prone to being inconsistently applied and defined, as well as containing too many opt-in initiatives, the bloc's response from 2018 to 2022 reflected a far tighter and multi-pronged approach (underlined by the mandatory DSA). The EU's response emphasises a need for collaboration, while also acknowledging the need for a degree of customisation built into codes of conduct (given the myriad ways that different platforms work). With regard to the 2021 News Media Bargaining Code (currently under review), there

are suggestions that Australia will eventually follow the EU's path, by placing extra scrutiny on platforms yet to sign up to the code (given the opt-in options).

- Australia adopted a slightly different approach during its 2022 election, whereby the Australian Electoral Commission (at federal level) were proactive in dealing with disinformation before it became a problem. This mirrored Narendra Modi's agreement with big tech platforms during the 2019 Indian election. After pre-empting problems with disinformation (including fake news about unsecure ballot papers, unreliable counting software and that unvaccinated voters would not have their votes counted), the AEC formed the Electoral Integrity and Communications Branch, using existing provisions within the Electoral Act to justify takedowns (as opposed to declaring themselves the arbiters of truth).
- Additionally, the AEC Statement of Intent set out how the AEC and tech companies would work together to address breaches of electoral laws and tech companies' terms of service, demonstrating a multi-pronged approach. The AEC were very active on social media, adopting memes and humour, which can be an effective tool against misinformation.
- However, it is important to acknowledge that not all misinformation shared and amplified on social media and online spaces is deliberate or malicious in nature. Furthermore, within big tech platforms' regulatory frameworks, there is an ongoing tension between universalism and customisation. While some MNCs have adopted a one-size fits all approach, there is instead a need to be flexible and acknowledge different cultural contexts. This would also lead to an educative and dialogic approach to regulation- for example, one that engages indigenous Australians, and raises greater public awareness toward their rights.
- Ultimately, a whole of society, multi-pronged approach is required. This needs to involve education (media and digital literacy), engagement and collaboration with all sectors (policymakers, tech companies, etc.) collaboration between different tech companies, greater transparency, and empowering public journalism (this would involve balancing civil sovereignty with avoiding chill effects and maintaining a vibrant public sphere).

SYNDICATE ROOM DISCUSSIONS RELATED TO PANEL 3

ISSUES: Engagement between platforms and policymakers

Multiprong approaches that engage and collaborate with multiple stakeholders are required to address the challenges in platform regulation. It is in the interest of platforms to ensure the freedom of expression and safety of their users, while maintaining revenue streams. Despite similarities within platform content and mechanisms, platforms themselves are distinct, and operate under different regulations depending on their respective regions of operation. Different social media platforms also maintain different levels of engagement with governments.

Twitter and Meta have been coordinating with policymakers unlike other platforms such as Weibo and TikTok. Media pressure can be put on platforms in order to prompt meaningful action. Some countries have opted to apply pressure by threatening to curtail operations within a country if certain safety safeguards demanded by the government are not followed. However, the implementation of regulations and fines can also result in retaliation by platforms. For example, Facebook threatened to block sharing of news content on its platform in Canada over concerns about legislation that would compel digital platforms to pay news publishers.

ISSUES: Role of civil society groups in the fight against misinformation

Civil society groups can help with content moderation and promoting media literacy. Local engagement by platforms with NGOs or policymakers is required in order to better understand local contexts and nuances. It is important for platforms to build alliances civil society groups- these alliances should maintain enough critical distance to support meaningful collaboration, but also to provide constructive criticism. In Southeast Asia, there is an added complexity of varying degrees of freedom of expression and cultural norms across (and within) states, which makes resilient collaboration between civil society groups and governments difficult- further highlighting problems with a one-size fits all approach.

Issues: Regulation of tech platforms

Due to a lack of commonly-accepted definitions and norms, platforms bypass and evade accountability due to difficulties in attributing intention, instead using terms such as "inauthentic behavior". Ultimately, platform regulation must be carried out at a design-level in order to prevent eventual monopolization. A preferred model would be co-regulation, whereby platforms are checked by governments, and an oversight board for self-checking is present within both the government and platform in question.

Panel 4: Strategic Communication and Alternative Media Channels

Experiences in Countermeasures

Colonel Jason Wright, *R/GEC Senior Military Advisor, Colonel, US Army, Global Engagement Center (GEC), Department of State*

- This presentation outlined the role of the Global Engagement Center (GEC), which forms part of the US State Department.
- The GEC has a clear mandate to study and counter misinformation, disinformation and propaganda from both foreign and non-state actors via various State Department platforms and allies. Mechanisms that are used to spread disinformation and propaganda are studied via multiple data-driven channels such as research and technology outreach, as well as tools to assess, evaluate and expose.
- The GEC's main objective is to work pro-actively with allies and partners to establish resilience. Separate specialised divisions for China, Iran and Russia have been established to work with regional bureaus and in-country embassy teams. Each of these divisions have country-specific goals, as well as separate lines of effort to analyse data and co-ordinate US government responses.
- The GEC sponsors research from external organisations, in addition to conducting their own research with over 40 in-house data scientists. The output from their work is then shared either publicly or via partnerships.
- The GEC also has a Technology Engagement Team that runs multiple activities in partnership with big corporations, including talks. A notable recent example includes a talk organised in partnership with Microsoft, on propaganda originating from the People's Republic of China. This team has also established an online gaming platform as part of their outreach and awareness building activities.

- Finally, the GEC Analytics & Research team work on data-driven projects using scraping tools and deploying both quantitative and qualitative approaches. Of special importance is GEC-IQ Global, an analytics and information sharing platform that other governments around the world can avail of to access and share information.

A Fake New World: Fighting Mis/Disinformation with AI, Edutainment and Social Media Literacy

Priyank Mathur, *Founder and CEO, Mythos Labs*

- Mythos Labs is a private company that uses media and technology to combat mis/disinformation, violent extremism, and online harms.
- Misinformation has existed for centuries with many examples from history. Modern-day misinformation, however, has since shifted to the online world, fuelled by the rapid uptake of mobile technology and long hours spent online on social media.
- There are three common disinformation techniques widely deployed and worth inoculating society against:
    - The first is *spoofing*, whereby false headlines are created by imitating the look and feel of reliable online news sources.
    - The second is *truthing*, a powerful technique whereby mainstream narratives are debunked with supposedly 'real' facts.
    - Finally, *decontextualizing*, where the truth is taken out of context.
- There are also two new technologies that bad actors exploit in order to spread disinformation:
    - The first is AI-powered content generation using tools such as DALL-E 2 and OpenAI GPT-3. These tools can be used to quickly produce large amounts of disinformation and manipulated content for a modest fee.
    - Another is the decentralized metaverse, in which audiences experience disinformation in a multisensory manner. In this metaverse, actual alternate realities are created, beyond the simple storytelling that is more prevalent on today's Internet. Crucially, there is, at present, no central authority on the metaverse to guarantee users' safety.
- There are three potential ways to address the above techniques and technologies, namely using AI, improving social media literacy and fighting viral lies with viral truths:
    - AI can be used to identify, monitor, and disrupt disinformation networks, and is of great interest to the military. As influence operations have become increasingly integrated with military operations worldwide, AI-based tools can be used to identify bots using machine learning and network modelling, and determine whether or not these bots are indeed acting in a co-ordinated manner, as well as decipher the disinformation they spread.
    - Social media literacy skills are a 'vaccine' for the misinformation virus. These skills counter polarisation and its real-world effects, including violent radicalisation. Social media literacy skills are especially crucial for users in developing nations, who lack basic knowledge about communications technology- for example, knowing how to detect and block radicalisation messages on social media- and are as such especially prone to online radicalisation.
    - Efforts to fight viral lies with viral truths involve using edutainment and data-driven counter-narratives. In a partnership between local NGOs and social influencer comedians, AI programmes were trained to communicate like the influencers by analysing all their historically-posted tweets, and generating new tweets. A focus

group subsequently deemed the AI-generated tweets to be more popular than the original influencer tweets.

- Combining social media literacy and edutainment is Mythos Lab's SMILE (Social Media and Internet Literacy E-module). This module is aimed at teenage Indian internet users who are frequent targets for misinformation and propaganda, and who are most likely to spread such content regardless of its credibility. SMILE has received positive feedback and produced encouraging results in schools.

Networks, Openness and Innovation – the UK's Communications During Russia's Invasion of Ukraine

Henry Collis, *Deputy Director, National Security Communications Team, UK Cabinet Office*

- The spread of disinformation during crises and politically-sensitive periods is by no means a new challenge- this phenomenon was particularly observable in England during the time of King James I, as well as during the Cold War. Crucially however, the UK government's approach to counter disinformation has evolved over the years. There have been several factors that have informed this evolution.
- While it should be acknowledged that there are a range of threats, state or otherwise, this presentation focused on the role of state actors.
- The UK first noted the threat posed by the Russian state following its illegal annexation of the Crimea in 2014. That same year saw the shooting of Malaysian Airlines Flight 17, as subsequent distractions, distortions and dismissals led to the obfuscation of the incident. The government subsequently recognised a pressing need to understand the Russian disinformation playbook.
- The UK is aiming to expose Russian tactics as a whole, and educate the public on these tactics, as opposed to merely dissecting individual incidents or campaigns. It has identified three components of the threat posed by disinformation to national security and society: the identity of the actors spreading disinformation, the environment in which the disinformation had taken hold, and the vulnerabilities of the audience.
- In order to effectively address the above components, the UK government first had to develop its own capabilities in identifying, assessing and countering disinformation- including ensuring that wider media and online spaces do not promote disinformation- and raising public awareness about the dangers posed by disinformation.
- The UK government devised its RESIST model with the objective of reducing the impact of disinformation campaigns on British society and national interests, by equipping the public with confidence in their own ability to assess the veracity of information themselves. Another similar initiative-since retired- was the SHARE campaign, which aimed to increase the resilience of the British public to disinformation campaigns, especially among those aged between 18–34.
- The UK government have also identified three areas of intervention, namely networks, openness and innovation. On the networks front, there has been a significant emphasis placed on interactions and building links between government and civil society. On openness, information from British platforms has been shared with allies and partners to help reinforce the UK's resilience to disinformation, by bringing collective voices together against Russia.
- British intelligence started becoming more public in November 2021 amid the build-up of Russian troops near the Ukrainian border. While the government were initially reluctant about sharing confidential information- particularly amid public scepticism in the aftermath of the Iraq War- it also acknowledges that intelligence is now being used for a completely different purpose, namely, to prevent a war instead of starting one. There has been significant

innovation in the way that this content is organised, as well as the pace and volume of its release.

- While the UK government's efforts to counter disinformation ultimately failed to deter the Russian invasion of Ukraine, they nevertheless exposed the false pretexts behind the invasion, pointing toward open-source evidence of Russia's body of falsehoods, which was then publicly shared. An example of this was a Sunday Telegraph article outlining the UK's belief that Vladimir Putin had planned to install a puppet regime in Ukraine, and identifying the possible puppet replacements. These articles have generated a strong resonance with Western audiences, albeit one that seems to be declining as the war continues.

SYNDICATE ROOM DISCUSSIONS RELATED TO PANEL 4

ISSUE: Challenges in framing counternarratives

Some methods used to frame counternarratives use locally-relevant ways to reach local audiences, by deploying strategic communications at a grassroots level. For instance, dialogues with Russian speaking minority groups in Estonia have led to a greater understanding of and engagement with Estonia's stance against the Russian invasion of Ukraine. This has also provided minority groups with an assurance of full representation within the Estonian national identity. Nevertheless, recent Russian information campaigns have been utilising narratives such as "colonial legacies" and "new imperialism by Western powers", and weaponizing such themes in support of Russian objectives.

ISSUE: The role of media literacy in countering misinformation

Billing media literacy as a cure-all is premature. There is first a need to distinguish between critical thinking and media literacy. Media literacy can involve the process of educating people on platform usage; for instance, walking users through the mechanics of blocking trolls and disabling accounts. It cannot be taken for granted or assumed that all users know how to protect their safety online. Educating individuals on the mechanics of keeping safe online is therefore highly crucial.

Governments are also working alongside the media to improve media literacy. The UK government, for instance, has been working closely with media platforms to counter misinformation. The aftermath of the annexation of Crimea saw the UK government working with the media to counter Russian misinformation. The UK government is also working with governments in Southeast Asia to share best practices in combatting misinformation.

ISSUE: Strategic communication strategies employed by governments

The UK government has previously worked with communication companies to aid in strategic communication. Such efforts include the engagement of individuals susceptible and/or vulnerable to terror recruitment. The US Global Engagement Centre (GEC) has placed a similar focus in counter-terrorism initiatives, including the introduction of a new character in the localised version of Sesame Street in the Middle East and North Africa. The repackaging of content can be further localised by influencers, comedians, and local content creators. Such strategies highlight the absurdity of a particular situation humorously in order to counter misinformation – and in turn, violent extremism.

ISSUE: Technological challenges in countering misinformation

There has been an increase in the offensive use of bots to supplement information campaigns. Some private sector companies have also pitched to governments the use of AI controlled bots to counter

misinformation. Governments, however, remain cautious of the use of bots, as their inauthentic behaviour does not align with democratic values upheld by many countries. Additionally, due to the risk of identity exposure, governments do not find such tactics worth utilizing.

New technology, such as the decentralised web and metaverse, poses another threat in the fight against misinformation. Governments, academics and NGOs must anticipate how misinformation and malicious activity would likely occur in this new space. These include avatar assault, or the creation of virtual spaces for hate speech in the metaverse. We need to study the legal and moral implications, and whether the metaverse would exacerbate risks. There are no legal frameworks to govern or regulate such spaces, so we should establish dedicated teams to investigate metaverse regulation.

ISSUE: Disinformation and countermeasures in diasporic populations

Although strategies designed to combat disinformation and misinformation share some similarities with those aimed at countering violence extremism (CVE), CVE strategies cannot be transposed wholesale due to subtle but important differences- such as the decentralized and unstructured structures of some information campaign actors.

Aside from monitoring and tracking, disinformation narratives can also be countered through dialogue. Albeit a lengthy process, it provides the opportunity to reduce the exploitation of fault lines in society.

Russian disinformation about Ukraine was initially targeted and directed at English-speaking audiences in the West (i.e., the US and Europe). Such attempts were ultimately unsuccessful in gaining traction among Western audiences. However, Russian tactics were subsequently switched to target disinformation at non-Western countries and allies (for example, in Asia and beyond), highlighting the importance of countermeasures in these countries in addition to the West.

Panel 5: Big Corporations and the Future of the Internet

Hard Coding Rights Into the Internet: What's Needed to Architect a Resilient Information Ecosystem

Quinn McKew, *Executive Director, Article 19*

- This presentation outlined the role and activities of Article 19- a 'think-do tank' with an emphasis on human rights and promoting freedom of expression.
- One of Article 19's areas of focus is the architecture of the Internet, which has a profound impact on the freedom of expression online. This architecture consists of three layers:
    - Physical infrastructure, including underlying telecommunications networks, Internet backbones and computing power
    - The logical layer, which includes root servers and domain names, which come under the purview of standards bodies such as ICANN, 3GPP and IETF
    - The economic and societal layer, which most internet users are aware of as it consists of the content they consume, including entertainment, social media and education- as well as the laws, policies and regulations that govern this content
- While Article 19's work covers each of these layers, it is the logical layer, consisting of standards and protocols, that forms the basis for the Internet and can therefore have a profound impact on freedom of expression. There is a significant role for human rights organisations in shaping standards and protocols, as well as a concerted push for public service technologists to work with these standards bodies in order to build upon the significant improvements that have been made so far, including a rehaul of the by-laws at ICANN and a new IETF protocol on human rights considerations.
- However, a point of serious concern is the push by certain countries to switch the governance of standards and protocols from a multistakeholder model to the International Telecommunication Union (ITU) multilateral model. This is concerning as the proposed ITU model may allow for content moderation and censorship at the Domain Name Server (DNS) level. This would easily allow the removal of complete websites from a country's internet, enabling a form of censorship.
- Moreover, in addition to the economic and societal layer of the Internet, disinformation can now also take hold in the logical layer. Internet infrastructure is itself now fast becoming a target of disinformation attacks. An example of internet infrastructure that can be attacked is digital certificates, which are used to encrypt online data flow between an end-user's browser and a website. Iranian hackers have reportedly breached a Dutch certificate authority to create fake Google certificates that would allow them to spy on Iranian Gmail accounts. Another example of disinformation in infrastructure is attacks on the Internet of Things (IoT) devices, harming their users and critical systems, and bringing disinformation to a whole new level. Such attacks on infrastructure that result in voter suppression may be an easier strategy than manipulating information during elections.
- The battles for control over standardisation at the ITU are another matter of concern. For instance, the Chinese government, along with Chinese companies, are hoping to set AI and biometrics standards that embed surveillance features and user behaviour analysis in IoT devices such as smart streetlights. There is also a drive to normalise the recognition of users' emotions, which can be an invasion of privacy.
- There is also lack of capacity building and bandwidth between the ITU and individual nations. In particular, many African and Asian nations have adopted ITU standards as their own without due consideration of their human rights implications.

- In conclusion, every technical standards decision is an opportunity for organisations to make the right choice in design and deployment. Infrastructure and logical layer services must also conduct human rights impact assessments when finalising standards and protocols.

How Online Misinformation Affects Brands and Young People, and Solutions for Mitigating These Threats

Veena McCoole, *VP, Strategic Partnerships, NewsGuard*

- NewsGuard provides independent ratings for over 8000 news and information websites. Its ratings are conducted by trained journalists, as opposed to using AI.
- While some news sites may appear reliable at first glance, NewsGuard reviews have revealed certain sites to have failed to meet basic editorial and journalistic integrity standards- for example, failures to disclose the sources of the site's funding. NewsGuard provides "Nutrition Labels" showing and explaining to users why certain sites have failed to meet these standards.
- To create these ratings and Nutrition Labels, NewsGuard works with partners including Internet companies, advertisers and AdTech companies, education service providers, governmental organisations, non-profit organisations, and researchers.
- NewsGuard helps advertisers to comply with online harms standards established by the Global Alliance for Responsible Media and the EU Commission. When advertisements from major brands appear on false news websites, these brands and advertisers risk undermining their customers' trust, and are inadvertently funding disinformation websites- when they could in fact be funding good quality journalism instead. NewsGuard helps advertisers to avoid these types of websites, in order for them to protect their own business reputations, and for their advertisements to generate better returns on investment. As advertising companies want to generate better profits without inhibiting their reach or limiting advertising opportunities, NewsGuard helps these companies to reach this goal.
- NewsGuard is also very concerned about the disinformation risks that young people face when navigating social media platforms like Tik Tok to consume content. This is because content on TikTok that contains misinformation tends to get more engagement. NewsGuard have also noted a trend among young people to using Tiktok as a search engine, which is worrying given that TikTok users are consistently shown false and misleading claims upon searching for information about prominent news topics. For example, in their first 35 minutes on TikTok, nearly 80% of users are shown misinformation related to COVID-19. Additionally, a NewsGuard investigation found that misinformation purveyors were able to use rudimentary techniques to bypass TikTok's moderation systems and post dangerous herbal abortion content, despite the platform's pledge to crack down on such content.
- NewsGuard believes that a human-centred (as opposed to AI-based) approach by trained journalists is most the effective way to combat misinformation on websites and social media platforms. Advertisers must also pay heed to where they advertise so as to ensure they are not funding misinformation. Finally, it is important for social media platforms to improve their own user tools, so that their users are better empowered to make the right decisions when consuming content.

Collaboration Between Civil Society and Private Sector

Alice Budisatrijo, *APAC Head of Misinformation Policy, Meta*

- This presentation outlined Meta's strategy in dealing with misinformation and disinformation, both defined as follows:
    - Misinformation is false or misleading content that is often shared unintentionally
    - Disinformation involves the deliberate intent to mislead or manipulate
- Meta has a three-prong strategy to deal with misinformation and disinformation. The first part of this strategy consists of the removal of content that violates community standards on hate speech, spam, fake accounts and including harmful misinformation. However, not all misinformation is removed. Meta adopts a selective approach to content removal, in order to avoid assuming the role of the arbiter of truth with regard to any and all information- such a role is neither enforceable nor feasible for any social platform.
- However, Meta maintains a clear mandate to remove misinformation that poses an imminent risk to physical harm, such as vaccine misinformation that could contribute to vaccine hesitancy, manipulated videos, deepfakes (created by AI) and voter suppression. At the time of this presentation, 25 million pieces of COVID-19 misinformation had been removed. Between April and Jun 2022, 1.4 billion fake accounts- representing 5% of global monthly active users on Facebook- were also taken down.
- Meta also maintains strict voter interference policies on Facebook. These policies disallow misrepresentation of key voting logistics information, such as dates, locations, times, methods for voting or registration, eligible voters, voting criteria, the eligibility of votes, as well as information about candidates.
- Meta disallows six types of behaviour during elections, including offers to buy or sell votes, distributing material that promotes illegal participation in voting, claims that voting would lead to COVID-19-related or law enforcement consequences, mass calls for interference that would suppress voters, and calls to monitor elections combined with the possibilities of violence.
- The second part of Meta's strategy focuses on reducing of distribution of low-quality content (including false news). This includes false or altered stories, Facebook Pages or domains that repeatedly share false news, spam, and or sensational content. Meta currently works with 80 third-party, certified fact checkers in 60 languages across 17 countries in Asia-Pacific on this strategy.
- The third part of Meta's strategy involves informing users about the content they are viewing or about to repost. This includes pertinent information about content- for example, content that has been flagged by an independent fact-checker as misleading- contextual references, and if the content is old or outdated. Meta have also launched a flagship programme called "We Think Digital", designed to improve users' digital literacy.
- In conclusion, while misinformation might increase user engagement in the short term, Meta is focused on the long-term health of its social platforms. Meta emphasises a zero-tolerance approach toward misinformation and disinformation, as part of their aim to persuade users to return to their platforms.

SYNDICATE ROOM DISCUSSIONS RELATED TO PANEL 5

ISSUE: The pace of technological developments in the fight against misinformation

While technological development will form part of the solution against online falsehoods, datasets presently going into AI continue to reflect human biases, making it difficult for AI to detect new forms of misinformation. Furthermore, as AI becomes increasingly complex, tools which were once regarded the solution have the potential to evolve into problems themselves. A continued stalemate between technology and human users is likely. At some level, the problem lies in the way humans receive information. It is still uncertain if we are able to embed human rights principles in fundamental internet architecture, especially if this is decentralised. The potential rise of a decentralised 'creator economy' would also weaken the credibility of established institutions, suggesting worrying changes for the institution of journalism as a whole.

Many current solutions are reactionary amidst the changing media landscape- an example is the UK Online Safety Bill (which has a trade-off between free speech and a heavy-handed government approach). Initiatives such as the EU Code of Practice on Disinformation, while trying and slow-moving, are more ideally conceptualised. It is up to platforms and civil society groups to appropriate the right solutions.

Policymakers and politicians are still lacking in their understanding of technology. This places big tech companies in an incredibly powerful situation, as policymakers are therefore willing to place full faith in companies to develop solutions on their behalf- hence leading to very bad solutions.

ISSUE: Differences in international and national policy-based solutions for countering misinformation

The EU policymaking landscape is fundamentally different from that of the US. On top of significant polarisation in the US, American legislation tends to centre around the concepts of freedom of expression and excessive moderation, leading to a contest between pro and anti-regulation advocates (in turn leading to very different policy outcomes). In contrast, European regulations standards are perceived as benchmarks- European policymakers have positioned themselves as the vanguard of a different set of policies such as the Digital Services Act. While the EU champions a multistakeholder approach, encouraging capacity-building between government, civil society, and academic institutions; there is currently a lack of accountability within the EU Code of Practice in its voluntary stage, making it difficult to effect real and meaningful change.

Heavy censorship is problematic from a human rights perspective. The majority of prosecutions under fake news laws have targeted legitimate news sources that published information deemed politically inconvenient. Questions remain over the intentions of such laws, as well as the quality of implementation. There are many problems with this approach that are yet to be addressed. There should be a clearly-defined emphasis on creating information environment supportive of democracy and human rights. There are many ways the internet can be used to lead to a disempowered citizenry and subjugate institutions.

There has also been advocacy for power to be invested in state mechanisms. While bodies such as the Internet Engineering Task Force are likely to remain autonomous, there have been attempts by Russia and China to co-opt bodies under UN auspices. This is part of a wider trend toward increasing local

sovereignty and control over the Internet. Such a brute force mechanism of exerting control is deeply problematic for a web based on democracy and human rights.

ISSUE: Considerations for code of conduct and resilience in Web 3.0

User empowerment tools are crucial in the transition from Web 2.0 to 3.0, and form part of platforms' duty of care. These 'better for you' empowerment tools and principles can be taken into the gaming world and in the context of Web 3.0. While Meta has a code of conduct for virtual experiences that covers harassment and violence, overall governance frameworks of the Metaverse are still in their infancy. Content moderation is likely to be carried out by users in their respective (online) spaces- as is the case in real life. Although platforms will not necessarily be stepping in to decide what users can and cannot say, content governance will ensure that virtual experiences mirror real-life experiences. Even as users are being driven and self-selecting into smaller communities, there is a still a need to safeguard their experiences. It is worth considering if the 'marketplace of ideas' on the Internet may become the fiefdoms of oligarchs, once competitive market pressures and regulatory pressures are absent. This would also worsen the existing problem of disenfranchisement in closed spaces.

ISSUE: Challenges for fact-checkers and media literacy initiatives

Despite the use of fact-checking, fact-checkers themselves often remain open to pre-existing biases and doubts (made especially clear during COVID). Fact-checkers may also have concerns about working in unsafe environments (e.g., in Malaysia or the Philippines).

Media literacy initiatives are also likely to be negatively affected by weakened trust levels. As governments have imposed firmer regulations in response to users continuing to engage in unsafe activity despite these educational initiatives, this has the potential to sow further distrust between platforms and governments.

ISSUE: Platform regulation and platform design consideration for safeguarding users

Publicly-traded companies have a strong incentive for immediate returns- as such, short-termism inevitably trumps longer-term business interests. Ultimately however, these platforms should prioritise long-term incentives instead. It is also important to note that advertising companies are reluctant to have their brands appear next to inflammatory content, which is itself a long-term concern for these companies. Inflammatory content designed to excite and engage users will do so on platforms that amplify such content. While clickbait headlines were initially very prevalent online, machine learning systems were eventually trained to recognise and demote these headlines. The financial incentives for clickbait and listicle articles eventually disappeared, and as such these websites no longer thrive.

As it is very difficult for governments and platforms to impose and enforce rules, it is crucial to empower users to check and discern sources on their own By instilling trust in the media via transparency, users can be made aware of the metrics to evaluate news sources using tools such as NewsGuard which emphasize transparency as opposed to partisanship, and subsequently feel empowered to exercise their own personal judgement.

Panel 6: Identity and Online Harms

Gender and Security in Digital Space

Dr Gulizar Haciyakupoglu, *Research Fellow, CENS, RSIS, NTU*

- This presentation focused on the gendered impacts of online harms and threats. While many scholars and practitioners explore a wide variety of harms from a security lens, there have been limited efforts to understand their gendered implications. The edited volume that was discussed, titled, 'Gender and Security in Digital Space: Navigating Access, Harassment and Disinformation' addressed this gap with a special focus on four areas 1) access 2) harassment 3) disinformation and 4) countermeasures.
- Digital spaces provide ample opportunities to advance gender equality in access to social, economic, and political venues and prospects. However, the same spaces also accommodate the inequalities, discrimination and harms that are prevalent offline, and they impinge on the Internet's aid to advancing gender equality and gender-sensitive security.
- There is a persevering access divide – not only regarding the Internet and device access, but also in terms of access to safe and informed online experiences. This is reflected by 2020 data from the International Telecommunication Union (ITU), whereby just 57% of women worldwide had Internet access, compared to 62% of men. This gap is even wider in Southeast Asia where, "women are 36 per cent less likely to use the internet than men". While the COVID-19 pandemic changed this dynamic slightly, inequality remains stark. Access limitations can also involve Internet shutdowns, which affect women disproportionately, creating an environment conducive to the regression of women's and social equality. For example, during Internet shutdowns in Kashmir, communication blockage prevented women's rights activists from offering counselling to women battling domestic violence.
- Access to skills and tech industries is another area where gender-based inequalities are visible. As per a 2022 World Economic Forum piece, new jobs will be driven by new technologies. This is a crucial indicator that the access gap and inequality should be closed as soon as possible to avoid further divisions, and that it is critical for women to have technical skills and presence in the industry. However, the same report also highlights that fewer than 15% of ICT professionals in G20 countries are women. Unequal representation in ICT and tech-related fields could also lead to ignorance toward some gendered threats and, at times, entirely wrong conceptions of vulnerability. Current victim and vulnerability definitions must be revisited to ensure safe and secure access. More importantly, access must come with safety and security, including from online harms such as harassment and disinformation.
- Gender-based online violence, including harassment, is wide-reaching and involves various malicious acts. In an Economist Intelligence Unit study, 85% of respondents witnessed online violence against other women, while 65% of women knew of other women from their professional or personal networks being targeted online. Observed threats included misinformation and defamation, cyber-harassment, hate speech, impersonation, hacking and stalking, Astroturfing, video-and image-based abuse, doxing, and violent threats. It is also important to acknowledge that these threats can be used in combination. The study also found that 88% of women in Asia Pacific had faced online violence.
- In the edited volume, discussions on harassment focused on dynamics in Southeast Asia, with examples and cases from Indonesia, Malaysia, and Singapore. Insights into the digital conditions in ASEAN member states provided striking examples of online gender-based violence. These include online human trafficking, forcing women and girls into unwanted

labour, marriage, and sexual engagements, and Malaysia's LGBTIQ+ community facing rape and sexual exploitation risks while using dating apps.

- Limitations on spaces for political speech may create further insecurities for gender equality advocates in Malaysia, especially when online harms challenge civil and secure online communication. One of the volume's sections concerned the scholarly arguments about using the 'Asian values' narrative to justify limitations of political speech, raising additional questions about gender equality advocacy. However, framing feminism as foreign, western, un-Islamic, or as a concept antagonistic to Asian values could lead to the sidelining of gender equality concerns and advocacy efforts, and cast gender equality advocates as disconnected from Asian society. This might create backlash and insecurities, or disrupt deliberation efforts.
- Another notable online harm with a gendered impact is disinformation. Some scholars and practitioners use the term 'gendered disinformation' in reference to false or misleading information, as well as visuals that build on preexisting gender stereotypes. Scholars have situated gendered disinformation within gender-based violence and stressed upon the importance of understanding gendered disinformation. This helps to clarify how sexist narratives are weaponized to discourage female participation in political and public deliberation.
- Gendered disinformation is used by state and non-state actors, including extremist groups- such as the Russia-backed disinformation groups targeting Ukrainian politician Svitlana Zalishuck by posting her face on pornographic images. Scholars and practitioners argue that women in public-facing positions, such as journalists and politicians, or "people with intersecting inequality factors" (such as gender and disability) are more vulnerable to these threats. Gendered online harms, including harassment and disinformation, have wide-reaching impacts- including job losses, mental health problems, physical harm, and limitations to the use of online spaces, which exacerbate financial, educational, and social disadvantages. Such threats also harm democracy by discouraging women and LGBTIQ individuals from running for office or participating in the democratic process. This compounds gender inequality in political representation, as well as findings from a UN Women study regarding executive government positions, which state that gender equality in the highest positions of power will only be achieved in 130 years at the current rate of progress.
- Various efforts aiming to alleviate the gendered impacts of online harms and to breach access gaps have so far been insufficient. As such, the edited volume contains six overlapping recommendations. They are:
    - Refining and establishing a consensus on the definitions of terms including gender-based violence, online harms, disinformation, harassment, and others with particular attention to gender and security dimensions.
    - Improving women's access to decision-making and leadership roles.
    - Understanding the role of information literacy.
    - Increasing awareness-raising efforts on the gendered impacts of online threats – in addition to the ongoing efforts of civil society organisations working on these issues, there are emerging initiatives and organisations that are providing spaces for women to be more vocal.
    - Domestic and transnational collaboration to combat the threats of disinformation.
    - Expanding research and gender-focused data gathering efforts.

Peacebuilding in the Digital Age: A Gendered and Intersectional Approach

Norul Mohamed Rashid, *Regional Policy Advisor for Governance, Peace and Security UN Women Regional Office for Asia and the Pacific (Bangkok)*

- This talk focused on cyber-security and online harms from the perspective of peacebuilding at the UN by taking a gendered and intersectional approach, positioned within the Women, Peace, and Security (WPS) agenda of the UN. This agenda was established by the Security Council in a 2000 resolution called 1325, in response to gender-blind peace agreements' provisions that did not invite women to the table for peace negotiations and failed to address the post-conflict needs of women and girls – completely excluding 50% of the population. In the last 22 years with the WPS, there have been 10 security council resolutions on women, peace and security which address emerging security issues where women are engaged- including in APAC, a largely peaceful region barring a couple of exceptions.
- Peace and security issues, having evolved over the last few decades, go beyond armed conflict, and have increasingly emerged online and in digital spaces. As a consequence, member states, chiefly the US and UK, have been greatly concerned with increasing their cybersecurity capacity, impacting their relationships with their allies. In response, the UN Security Council and General Assembly have carried out work on cybersecurity and its consequences on international security. In June 2021, the UNSC held, for the very first time, a debate on the influence of cyber threats on international security. The debate outlined the positions of the P5 including China and Russia, and also considered other aspects including International Humanitarian Law, the right to self-defence, invasions and military actions and if these applied in cyber space. Although there was no resolution from this debate, there has since been action to narrow the gender divide, and provide meaningful female participation and leadership in decision-making processes pertaining to cybersecurity in the context of international security.
- There is a lot of data on the lack of parity and underrepresentation of women in the tech field. Some examples are as follows:
    - a) in Southeast Asia, women only make up 39% of technology students and 32% of the total tech industry workforce
    - b) In Artificial Intelligence (AI)-related fields: men make up 78% of AI professionals
    - c) Women in Law Enforcement: Women's leadership and representation in cybercrime unit are low.
- These numbers are important as they impact policymaking and the way our powered systems are designed. Crucially, to be in the position to make any kind of impact in the digital space requires a range of skillsets across different regions- women oftentimes lack access, or are not equipped to deal with the proliferation of issues.
- The UN is working closely with many women human rights defenders amid myriad online harms and threats currently targeting human rights activists. Women (and LGBTQIA+) human rights defenders, in particular, are heavily targeted in the APAC region. Tactics include cyberbullying, trolling, outing, doxing, and disinformation. Disinformation aiming to discredit women holding public positions prevents them from exercising functions, and is usually accompanied with misogyny and violent extremism online. Misogyny is integral to the ideology, political identity, and political economy of violent extremist groups. Gender stereotypes and misogynistic narratives are frequently exploited in propaganda and recruitment strategies. Numerous instances of hate speech targeting LGBTQIA+ groups and gender equality advocates have been reported, including a rise in online misogynistic hate-speech during COVID-19. In May 2020, there was a 100% increase in week on week in tweets

containing misogynistic references. COVID-19 has highlighted existing gender biases and exposed that women are more vulnerable than men, both online and offline.

- Hateful online rhetoric tends to transcend digital spaces, reflected by the rise of numerous cyber-facilitated crimes. For instance, social media, marriage and dating apps are used by transnational criminal networks to deceive and draw users into exploitative and often degrading situations, commonly under the pretext of promised work opportunities. There are also cases of workers being trafficked to conduct cyber-crimes, such as sex scams and cryptocurrency romance scams. Similarly, high-ranking politicians, journalists, human rights defenders, and activists face a high risk of data breaches and hacking, as revealed by the Pegasus project. Data breaches may result in doxing, extortion, and other severe risks for women activists, which in turn may lead to increased self-censorship.
- The responsibility to address these harms ultimately falls on states. Legislative and policy frameworks are currently ill-equipped to address gender harms and cyber-enabled crimes affecting women. Cybersecurity-related legislation and policies are rarely explicitly linked to or harmonised with laws for women's protection and rights, and are poorly conceptualised under international law. More state efforts are required to increase awareness of law enforcement and the judiciary on cyber-facilitated crimes, and to improve digital media companies' response mechanisms.
- AI-powered systems, although a fairly new area of research, are known to be discriminatory and contain inherent gender biases. Initial findings of a study done by the WPS shows that 1) algorithmic gender and racial biases disproportionately disadvantage women of colour 2) the digital gender gap, lack of sex-disaggregated, and gender-sensitive algorithm programming contributes to bias and inaccuracy issues and, 3) algorithms that enable filter bubbles/digital echo chambers, bots and deepfakes are facilitating the spread of disinformation, misogyny, and hate-speech. These factors pose a gendered security threat and must be addressed.
- The most important recommendation to combat and navigate these risks is to increase digital literacy and awareness of gendered online risks. This will equip activists and journalists with the ability to conduct online advocacy and activism safely. WPS has a e-module for navigating disinformation specially developed in partnership with Ridgeway Information Limited and Girl Security, that is freely available online in many languages. The module includes sessions on disinformation, gender-based online discrimination and harassment, online hate-speech and violent extremism and, safe cyber hygiene practices. Its objective is to empower communities of women in the APAC region through digital literacy training and a strengthened capacity to use social media. There also needs to be continued research on the gendered implications of AI on the implementation of WPS' agenda in Southeast Asia, focusing on cyber-resilience.

Indonesia's Lesson Learned in Addressing Online Gender-Based Violence

Fitriani, *PhD, Centre for Strategic and International Studies (CSIS) (Indonesia)*

- This presentation focused on the rise of online gender-based violence (OGBV) in Indonesia. Online gender-based violence in the context of this talk included violence against women and girls in digital contexts, that also occurs as a continuum often connected offline and encompasses many forms. The Centre for Strategic and International Studies (CSIS) in Indonesia have developed a national hate speech dashboard that captures and identifies online hate speech. Apart from analysing the narratives of hate speech, the dashboard supplements existing qualitative hate speech studies in Indonesia by providing quantifiable data portraying the volume and actors that share hate speech content. It also provides

chronological patterns of how online hate speech intensifies, and is crucial in developing a better description of the magnitude and gravity of hate speech online.

- Although Indonesia remains reluctant to address the power and inequality of gender-based violence, the dashboard has identified that victims of online hate speech are mainly individuals belonging to religious and political minority groups, and lower-class groups. Recent geopolitical developments that have led to the rise of ultranationalism also contribute to the increase in hateful rhetoric online, including comments that intersect with race and gender. The dashboard contains a collective violence monitoring tool that supplies data to mitigate risks, and also tracks the occurrence and casualties arising from these behaviours offline. There was a stark increase in online violence towards gender minorities during the COVID-19 pandemic in Indonesia – despite the fact that most cases of gender-based violence are unreported as victims struggle to come forward. Indonesia has limited legislation in place to protect those who report such violence- ensuring that the act of reporting places victims in greater risk of targeted and continued harassment.
- While the Internet as a tool has helped Indonesia advance welfare and economic developments, the digital space continues to be unsafe for gender minorities. 75% of OGBV occurrences in 2021 took the form of non-consensual dissemination of intimate images (NCII), while the remaining 25% were online-threats, doxing, cyber flashing, flaming, impersonation, morphing, outing, invasion of privacy, sexual harassment, damage to reputation, and phishing. The most common perpetrators fall into three main groups – religious, anti-feminist and nationalist groups. But within these groups, it was found that 81.9% of the perpetrators were people with close personal relationships with the victims, while 1.8% belonged to the same organisation or institution as the victims. Women human rights defenders, in particular, are more vulnerable to these attacks as they limit their advocacy and also serve as an intimidation tactic.
- Online gender-based violence have wide-reaching health, wellbeing and safety impacts that are both physical and psychological in nature. The social and political impacts of OBGV need to be addressed as well – victims suffer from ostracization and persecution within their communities that also adversely affect their mental health leading to depression, self-harm and in severe cases, injury and death. This fosters a very unsafe digital environment for women, leading to a wider gender gap, further limiting their access to the Internet. The limited legal framework of Indonesia regarding victim protection and access to justice also poses a big challenge toward addressing online gender-based violence. For instance, it is common for victims to be sued by their perpetrator for defamation, forcing the victim to relive their trauma with their harasser. This gap is exploited by offenders to escape consequences with law enforcement.
- CSIS has provided four main recommendations to address the above:
    - Apart from raising awareness about online harms for women, young girls, and members of the LGBTQIA+ community, victims also need to be made aware of their rights within the legal system.
    - It is important to train law enforcement officers to handle sensitive cases, such as sexual harassment and NCII. Law enforcement agencies must be able to investigate and prosecute offenders of digital violence without revictimizing the victims and placing themselves and their family in greater danger.
    - There is an urgent need to review and reevaluate gender equality policy and legislation.
    - Lastly, social media platforms must provide security and equity guidelines designed to protect vulnerable groups on the Internet.

SYNDICATE ROOM DISCUSSIONS RELATED TO PANEL 6

ISSUE: The importance of definitions, and understanding gaps in establishing a consensus on terminology when discussing gendered issues

While it is important to refine and establish a consensus on the definitions of key terms, there should be less of a focus placed on definitions per se, and more on including varied gender perspectives, contexts, and experiences. Most work around cybersecurity is gender-blind, and a continued lack of these perspectives is harmful. For instance, we often hear that AI systems are biased against women, mainly because these systems tend to be designed disproportionately by men, and do not reflect women's experiences and needs.

Similarly, in the national security realm, gender tends to be overlooked. Most violent ideologies use misogyny as a tool to indoctrinate people. Although women face misogyny in daily life, both and online and offline- including sexual harassment and stalking- such threats to individuals are still not equated to national security threats. Furthermore, women who assume leadership roles are usually forced into situations where they are sexualised, and deal with disproportionate amounts of scrutiny over their gender and capability. In this context, men in power are largely responsible for building digital spheres and defining spaces that exclude women.

ISSUE: Women policing their own behaviour online

There is no justification for posting hateful rhetoric online. Unfortunately, victim-shaming remains very common in society. With any case involving harassment, laws that apply offline ought to apply online as well. While sexual harassment victims have used social media for viral justice- usually due to the absence of legal recourse- this also places victims at the risk of revictimization. In the Southeast Asian region, the "Asian values" narrative has been used to justify harassment and misogyny, further alienating women who are currently dealing or have dealt with these experiences. An additional reason for women to share their negative experiences online involves solidarity. Sexualization is a form of silencing, and it can be isolating for women to speak up against harassment in real life.

The ongoing rise of authoritarianism and conservatism in the Asia-Pacific region has also led to the shrinking of civic spaces. This is problematic, considering the need for diversity in representation when topics about gender and security are being discussed, and especially given that these topics have the potential to affect everyone in all spheres of society.

ISSUE: On the sharp rise of 'incels' and toxic misogyny online

The precipitous increase in 'incel' (involuntary celibates, an online community of men hostile to women) activity and misogyny online can be attributed to COVID-19. The UN has collected data on rampant misogyny in online spaces during the COVID-19 lockdowns. For example, in India, groups identifying as Men's Rights Activists (MRA) have become very popular and operate overtly. MRA groups are vehemently anti-women, and use chauvinistic language to intimidate and harass women online. These behaviours stem from personal insecurities, disdain for women and inability to handle rejection – coupled with other socio-economic factors like unemployment, this pushes people toward online communities where they indulge in violent misogyny. This is a significant security threat, as these attitudes can translate offline. Feelings of disenfranchisement within society are a major factor as they push individuals to find and further exploit identity cleavages in the social hierarchy- women with intersecting identities (e.g., female and minorities) are often the victims here. These components,

along with the empowerment gained from user anonymity, contribute to the strength and growth of the manosphere.

ISSUE: Content moderation with gender in mind

There is a pressing need for digital platforms to establish a playbook and timeline to address harassment complaints. This is because civil society organisations do not have direct access to content, and therefore cannot carry out content moderation. The UN has recommended member states to provide legal recourses for civil society organisations, which will allow these organisations to directly connect with platforms to address hate speech. Some civil society organisations have also launched free online modules designed to improve digital literacy within society. However, as much as regular citizens are equipped and empowered with digital literacy tools, it is ultimately the state's responsibility to safeguard these citizens, both online and offline.

Workshop Programme

Venue: Garden Ballroom Foyer, Level 1

Wednesday, 2 November 2022			
0800–0900hrs	**Registration**		
	Venue	:	Garden Ballroom Foyer, Level 1
0900–0910hrs	**Workshop Welcome Remarks** by ***Shashi Jayakumar****, Head, Centre of Excellence for National Security (CENS), RSIS, NTU*		
0910–1000hrs	**Panel 1: Alternate Spaces in Information Operations**		
	Chair	:	***Teo Yi-Ling****, Senior Fellow, CENS, RSIS, NTU*
	Speakers	:	**Content Moderation on Encrypted Platforms** by ***Kiran Garimella,*** *Assistant Professor, Rutgers University*
			Information Operations on Wikipedia by ***Abhas Tripathi,*** *Disinformation Manager, Wikimedia Foundation*
1000–1030hrs	**Networking Break**		
1030-1100hrs	**Special Presentation: Influence and Coordination: Detecting Interesting Activity in social media** *by* ***Kathleen M. Carley,*** *Professor, Carnegie Mellon University*		
1100–1200hrs	**Panel 2: Countermeasures**		
	Chair	:	***Shashi Jayakumar****, Head, CENS, RSIS, NTU*
	Speakers		**Co-designing a Mobile-based Game to Improve Misinformation Resistance and Vaccine Knowledge in Uganda, Kenya, and Rwanda**
			by ***John Cook*****,** *Associate Professor, Monash University*
		:	**The Weaponised Fact-Check: False Flags and False Checks** *by* ***Andrew Moshirnia*****,** *Associate Professor, Monash University*
			Checking on Fact-Checking: What's Effective and What Can Backfire *by* ***Edson C. Tandoc Jr.****, Associate Professor, Wee Kim Wee School of Communication and Information, NTU*
1200–1330hrs	**Networking Lunch**		
1330–1430hrs	**Panel 3: Regulating Platforms**		
	Chair	:	***Benjamin Ang****, Senior Fellow, Deputy Head, CENS, RSIS, NTU*
	Speakers	:	**Regulating Social Media Platforms to Protect Human Rights in Southeast Asia: A Civil Society Perspective** *by* ***Sutawan Chanprasert****, Executive Director, DigitalReach*
			Mitigating Digital Disruption: Strategies from The Asia Pacific and A Case Study of Combatting Disinformation During

Australia's 2022 Election *by* ***Andrea Carson****, Professor of Political Communication, La Trobe University, Melbourne*

1430–1500hrs **Networking Break**

1500–1630hrs **Interactive Syndicate Discussions**

1630hrs **End of Day 1**

1830–2030hrs **Welcome Dinner for Speakers**

Thursday, 3 November 2022

0800–0900hrs **Registration**

Venue : Garden Ballroom Foyer, Level 1

0900–1000hrs **Panel 4: Strategic Communication and Alternative Media Channels**

Chair : ***Dr Gillian Goh****, Adjunct Senior Fellow, CENS, RSIS, NTU*

Speakers : **Experiences in Countermeasures** *by* ***COL Jason Wright,*** *R/GEC Senior Military Advisor, Colonel, US Army, Global Engagement Center, Department of State*

A Fake New World: Fighting Mis/disinformation with AI, Edutainment and Social Media Literacy *by* ***Priyank Mathur****, Founder and CEO, Mythos Lab*

Networks, Openness, and Innovation - the UK's communications during Russia's Illegal Invasion of Ukraine *by* ***Henry Collis****, Deputy Director, National Security Communications Team, UK Cabinet Office*

1000–1030hrs

1030–1130hrs **Panel 5: Big Corporations and the Future of the Internet**

Chair : ***Muhammad Faizal Bin Abdul Rahman,*** *Research Fellow, RSAP, RSIS, NTU*

Speakers : **Hard Coding Rights into The Internet: What's Needed to Architect A Resilient Information Ecosystem** *by* ***Quinn McKew****, Executive Director, Article19*

How Online Misinformation Affects Brands and Young People, And Solutions For Mitigating These Threats *by* ***Veena McCoole****, VP, Strategic Partnerships, NewsGuard*

Fighting Misinformation at Meta by ***Alice Badisatrijo,*** *APAC Head of Misinformation Policy, Meta*

1130–1300hrs **Networking Lunch**

1300–1400hrs **Panel 6: Identity and Online Harms**

Chair : ***Yasmine Wong****, Senior Analyst, CENS, RSIS, NTU*

Speakers : **Peacebuilding In the Digital Age: A Gendered and Intersectional Approach** *by* ***Norul Mohamed Rashid****, Policy*

	Advisor on Governance and Peace & Security for Asia and the Pacific, UN Women **Gender and Security in Digital Space** *by* ***Gulizar Haciyakupoglu****, Research Fellow, CENS, RSIS, NTU* **Indonesia's Lessons Learned in Addressing Online Gender-Based Violence** *by* ***Fitriani****, Cyber Security Project Lead / Senior Researcher, International Relations Department, Centre for Strategic and International Studies (CSIS) Indonesia.*
1400–1430hrs	**Networking Break**
1430–1600hrs	**Interactive Syndicate Discussions**
1600–1630hrs	**Closing Remarks** by ***Shashi Jayakumar*****,** Head, CENS, RSIS, NTU
1630hrs	**End of Workshop**
1830–2030hrs	**Closing Dinner (by Invitation Only)**

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 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. Besides fulltime 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 to develop comprehensive approaches to strategic thinking on issues related to security and stability in the Asia Pacific. For more information about RSIS, please visit www.rsis.edu.sg.