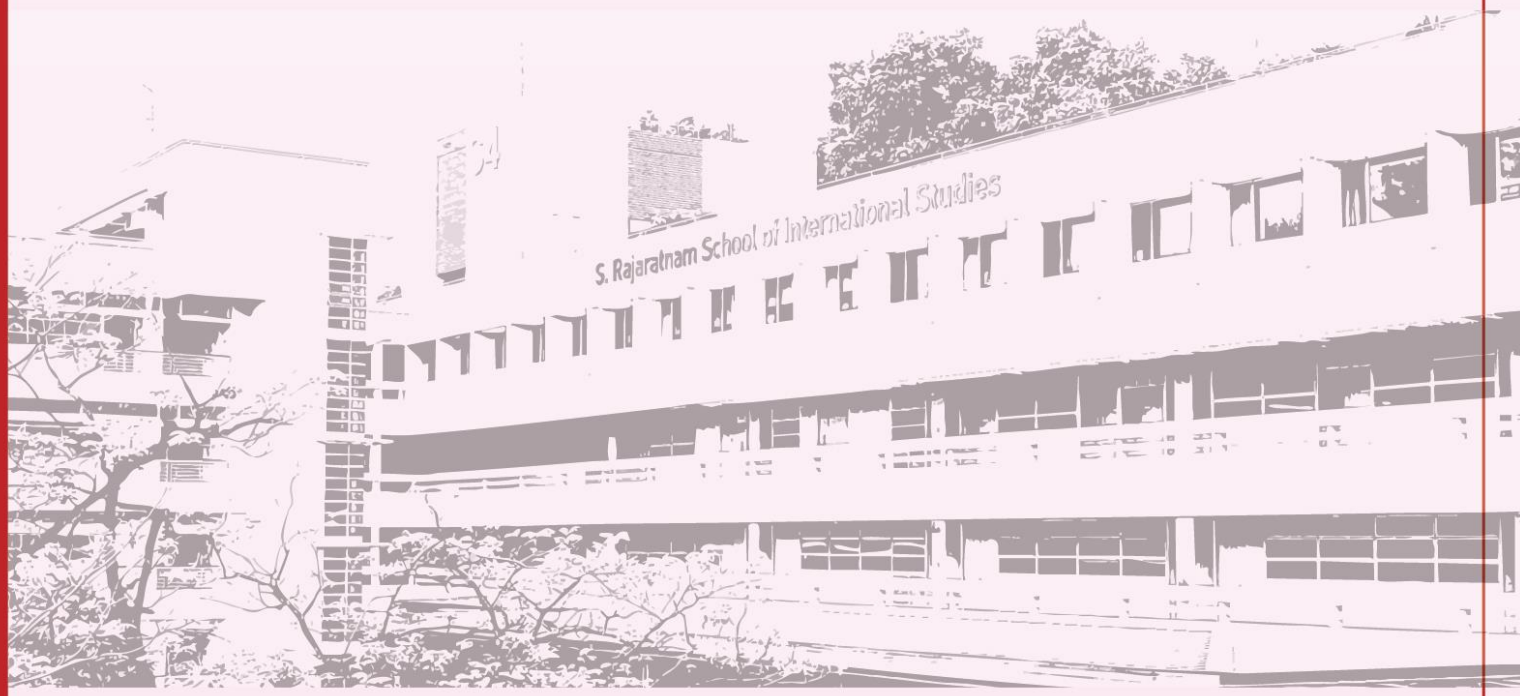
COMMEMORATIVE/EVENT REPORT MARCH 2025

Disinformation, Rumours, Untruths, Misinformation and Smears (DRUMS) Conference 2024



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

Ponder the Improbable since 1996

Contents

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: Tan E-Reng, Vincent Kyle D. Prada, Jermaine Soh, Dymples Leong, Antara Chakraborthy, Juan Cui Ying, Benjamin Chua, Yasmine Wong, Gulizar Haciyakupoglu, Sean Tan and Eugene Tan

Editors: Gulizar Haciyakupoglu and Dymples Leong

Terms of use: This publication may be reproduced electronically or in print, and used in discussions on radio, television and fora, with prior written permission obtained from RSIS and due credit given to the author(s) and RSIS. Please email to RSISPublications@ntu.edu.sg for further editorial queries.

Executive Summary

The Centre of Excellence for National Security (CENS) had its annual Disinformation, Rumours, Untruths, Misinformation and Smears (DRUMS) conference on 19-20 November 2024. The conference theme was "Information Manipulation and Interference in the Global Political Environment of Uncertainty".

Over two days, 180 participants from government agencies, academia, diplomatic corps, and non-governmental organisations learned from and engaged with 15 international and local speakers.

The keynote speech set the tone for the conference by exploring how the three converging areas of increasing systemic competition, rapid technological change and globalism will impact online information environments.

Panel 1 speakers spoke on information manipulation and interference. They touched on the exploitation of international conflict narratives by domestic influence actors in Europe and Southeast Asia. Panel 2 speakers spotlighted global cases of election interference and drew attention to the risks of disinformation narratives and generative artificial intelligence (Gen AI) in elections.

Speakers for Panel 3 elucidated the risks, uncertainties and the future of emerging technologies, including Gen AI. Speakers highlighted the importance of incorporating local/cultural nuances when exploring how experts can leverage Gen AI to combat disinformation and develop related technologies. Panel 4 speakers surfaced valuable insights from case studies in Malaysia, the United States and Vietnam. They stressed how preparedness, increased public awareness, and regulatory measures can play a part in safeguarding the online information space.

The 2024 edition of DRUMS received positive feedback on various aspects, including the design and organisation of panels, selection of speakers and topics, operation, and administrative matters.

The following sections of this report summarise key points from the speakers' presentations. Key takeaways from the Q&A sessions are available at the end of each panel.

Welcome Remarks

Ambassador Ong Keng Yong, Executive Deputy Chairman, S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University (NTU)

- The DRUMS (Distortions, Rumours, Untruths, Misinformation, Smears) conference has been an annual undertaking by the Centre of Excellence for National Security (CENS) at RSIS since 2017. DRUMS 2024 focuses on information manipulation and interference amidst uncertainties created by geopolitical rivalry amongst world powers. Armed conflicts and wars in various parts of the world have created a fertile ground for DRUMS activity, which has a significant impact on electoral processes and policymaking.
- Tactics, Techniques, and Procedures (TTPs) used to carry out DRUMS activity include utilising Information Technology (IT) to smear and spread untruths. This has necessitated a greater understanding of the psychology of misinformation and countermeasures to combat DRUMS lest they disrupt social harmony and public trust in governments.
- As misinformation can potentially influence voters, Singapore must address the threats posed by DRUMS, especially with a general election due to be called soon.
- In a 2024 report, the World Economic Forum listed misinformation and disinformation as the biggest short-term threats and reaffirmed the need to address their challenges.
- Malicious actors have used Gen AI tools to amplify disinformation in many countries, impersonating both politicians and celebrities. Existing laws and conventional practices can no longer contain these problems.
- DRUMS cannot be allowed to destroy businesses and time-honoured governmental practices. Emerging technologies and future uncertainties must be addressed through a whole-of-society effort and global cooperation.

Keynote Speech

Graham Brookie, Vice President and Senior Director, Digital Forensic Research Lab (DFRLab)

- During the 2016 elections, experts argued that Russia targeted the United States with large-scale Foreign Information Manipulation and Interference (FIMI) activity. These FIMI activities came at a time when the United States was not fully resilient against disinformation campaigns and was unprepared for a systemic confrontation between different governments of different political systems.
- The 2016 FIMI experience showed that strategic communication is a significant component of the policy portfolio. The current global competition for information has necessitated an understanding of how global information trends influence policy.
- In 2015, some researchers from the DFRLab investigated if it was possible to prove - using open-source information - if there were Russian regulars operating in Eastern Ukraine, when the Russian government denied having any Russian troops stationed there. The researchers proved the presence of Russian troops in the area by finding selfies posted by Russian soldiers on open social media platforms such as Instagram and VKontakte. Using geolocation with Google Earth, they confirmed the location and times of the selfies and correlated these data points to attacks in Eastern Ukraine.
- There are three converging trends experienced around the globe: (a) the increasing geopolitical contestation between the Great Powers and larger countries, (b) the rapid rate of technological development, particularly in the field of AI, and (c) globalism and widespread interdependence across the entire technological ecosystem.
- 2024 was the global year of elections, with more people going to the polls this year than any other year in recorded human history. In 2024, the number of global elections was roughly equivalent to the number of global elections in 2018 and 2020 combined. This volume is unlikely to be replicated till 2048.
- In 2024, governments were better prepared for information threats and vulnerabilities (including foreign influence efforts), particularly with regard to elections, which has built up a considerable amount of resilience. 2024 also saw increased economic volatility, which involved the dismissal of about 90,000 people working in trust and safety policy teams on larger social media platforms. Civil society actors in the disinformation space increasingly came under threat, with many financial resources supporting civil society having pulled out.
- 2024 saw the most aggressive policy responses against information operations in US history. The responses were enacted in real-time, unlike in past elections, when responses only came in after the elections.
    - Operation Doppelganger involved creating fake news websites fed with content to make them look like credible news outlets, where pro-Russian and anti-Ukrainian stories were posted. Influence operations with traces of traditional intelligence operations came in the form of online influencers being paid to espouse pro-Russian views.

- Proactive steps were undertaken to mitigate influence operations, along with policy responses such as the implementation of diplomatic measures and sanctions. However, influence operations continued, with actors such as Russia and Iran being the most active.
- China has also engaged in increasingly aggressive testing of new types of tactics in the information environment, as well as amplifying Russian operations. While there might not be a formal working relationship between the two powers, the overarching goals of increasing polarisation are aligned.
- From the immediate pre-election period to the post-election period, there was also a dramatic shift from broad influence operations to directed, targeted mobilisation for actions (e.g., encouraging abstaining from voting, or voting). Policy responses will have to demonstrate awareness of this threshold between broad influence and direct influence/mobilisation.
- There have been debates on how AI could impact the information environment, including during the election period.
    - The cases observed during the elections demonstrated that AI is pervasive, but not persuasive. To date, there have not been any cases where Gen AI has definitively changed or swung the results of an election. This has been partly attributed to the resilience of institutions and stakeholders. While Gen AI is used for a lot of things, including FIMI operations, audiences are not necessarily persuaded of a certain viewpoint as a direct result of AI.
    - Nonetheless, Gen AI accelerates the world's current trust deficit, creating an overall challenge for combating and mitigating information pollution.

Key Points Noted from the Q&A Session

Issue: The potential impact of a TikTok ban on the global information environment and the fragmentation of social media platforms along ideological lines.

- The argument that TikTok should be banned hinges on the assertion that TikTok is a systemic tool for the Chinese Communist Party (CCP) to assert its influence. However, given the context of an ongoing global competition for information, the core necessity to have a functioning democracy is an open information environment. When power is projected around the world, US interests are better served by having a better information environment. Hence, banning TikTok might not be beneficial. However, if there is substantial evidence that TikTok might be used as a vector for foreign interference, it would need to be mitigated. The law put in place by the CCP that requires citizens to turn over any information that might be of interest to it is commonly cited as evidence of TikTok's potential to be used as a vector for FIMI. Whether or not the CCP is using TikTok to obtain data about US citizens is largely irrelevant in the grand scheme of things as the same data can be bought (likely for cheaper) through data brokers, even if a TikTok ban were to be enforced. In addition to the TikTok debate, social media platforms are fragmented along ideological lines in the United States. Smaller platforms, especially after the 2020 US elections, have seen higher engagement.

Issue: Desensitisation and acceptance of FIMI and the possibility of government responses to FIMI backfiring.

- Experts have observed desensitisation in cases where foreign influence is framed as a subject of political debate. In responding to FIMI, authorities have a responsibility to be direct and transparent about the levels of threats being faced. They should present the nature of threats as they are without exaggerating or downplaying certain aspects.

Panel 1: Information Manipulation and Interference in Times of Uncertainty

Strategic Logic of Domestic Spreaders: Why Intra-European Networks Contribute to Foreign Information Manipulation and Interference (FIMI)?

Dr Akin Unver, Associate Professor, Ozyegin University

- Local misinformation spreaders within intra-European networks have enabled foreign interference, by embedding themselves within local ecosystems. and amplifying external influences within domestic online ecosystems. Foreign interference is rarely effective without local collaborators—whether political entities, media organisations, or online influencers—who lend credibility and reach to external narratives.
- Existing disinformation frameworks have been adapted from information warfare's "kill chain" concept. Frameworks such as DISARM (Disinformation Analysis and Response Mapping) provide a structured methodology for analysing adversarial tactics and implementing defensive countermeasures. DISARM has been widely utilised by governments and organisations to gain insights into strategies, vulnerabilities, and countermeasures to mitigate the impact of disinformation.
- Other frameworks included the DE-CONSPIRATOR project. DE-CONSPIRATOR maps and analyses information suppression efforts and potential transnational impacts in the European Union. The initiative underscores the importance of understanding the interplay between foreign actors (e.g., China and Russia) and domestic actors in enabling and amplifying disinformation campaigns.
- Key objectives of disinformation campaigns include influencing public opinion, destabilising governance and eroding trust. A comprehensive strategy to counter Foreign Information Manipulation and Interference (FIMI), which integrates technology, strategic frameworks, and public engagement, is crucial, alongside international cooperation and institutional transparency. Addressing external threats and internal vulnerabilities can help build and sustain long-term resilience in democratic systems.

"Glocalising" digital propaganda: How influence actors in Southeast Asia exploit geopolitical conflict narratives

Dr Janjira Sombatpoonsiri, Research Fellow, German Institute for Global Area Studies (GIGA) and Assistant Professor, Chulalongkorn University

- There have been cases of influence actors in Southeast Asia that have leveraged geopolitical narratives to shape public opinion and manipulate political discourse in countries such as Thailand, Malaysia, Indonesia, and the Philippines. The role of domestic actors is crucial in influence efforts within the region and is often aligned with or supported by foreign interests.

- Influence operations observed within the region are embedded within local contexts and are carried out through sophisticated mechanisms that include state-backed initiatives, institutionalised propaganda networks, political party sponsorships, and private entities disguised as independent organisations, such as quasi-think tanks and pseudo-media outlets.
- Various geopolitical narratives were tailored to exploit nationalistic sentiments and historical scepticism toward Western influence. In Thailand, for example, election campaigns have frequently featured anti-U.S. and pro-China rhetoric, framing the U.S. as an external meddler while portraying China as a benevolent partner. Protestors opposing government policies have often been stigmatised as "traitors" or accused of being influenced by Western agendas. These narratives discredit dissenting voices and create a polarising environment that discourages public debate and consensus.
- Influence operations thrive on pre-existing societal conditions. Local ecosystems, such as the political climate, media landscape, and historical relations with major powers, play a critical role in determining the receptivity of populations to foreign influence operations. For instance, in regions where nationalism intertwines with anti-Western sentiments, narratives promoting sovereignty and scepticism toward Western imperialism are particularly effective. Effective narratives include the framing of global conflicts, such as the portrayal of Ukraine as the provocateur in the Russia-Ukraine war, and narratives promoting pro-China and pro-Russia messaging strategies.
- Influence actors have been observed to deploy and leverage multifaceted strategies. Leveraging cultural products, social media campaigns, and alternative media channels to amplify their narratives and an understanding of local developments and culture has amplified the effectiveness of various disinformation narratives.
- Countermeasures to tackle influence operations involve addressing the domestic factors that make societies vulnerable to manipulation. These include fostering media literacy, promoting independent journalism, strengthening democratic institutions, and encouraging critical public engagement to build resilience against external and internal propaganda efforts.

Key Points Noted from the Q&A Session

Issue: Impact of then-US President Joe Biden's decision to allow Ukraine to strike Russia with long-range missiles on strategies pertaining to Russian disinformation.

There might not be major changes to Russian disinformation in Ukraine during the war. These main narratives include the threat of nuclear weapon deployment by Russia in attempts to deter allies of Ukraine from escalating the war further and to claim that actions by Ukraine and its allies would contribute to an outbreak of global war.

Issue: Divisive political issues (wedge issues) are among the tactics observed in information operations campaigns across countries in Southeast Asia and Europe.

The focus on the involvement of influencers and public relations companies in information operations has eclipsed the participation of volunteers in such activities.

Volunteers can be primarily driven by various ideological motivations or by their affinity with a political institution or a foreign actor. The volunteers can appear more authentic and convincing than other efforts to persuade potential voters to support the desired candidate. Influence campaigns in Europe depend on many factors, although ideological factors are an observable point to note in the context of elections in Europe. Far-right groups in the European Union member states can tap on domestic political struggles and understand the local nuances within a country.

Issue: Raising public awareness of FIMI without further fuelling divisive sentiments held by a part of the population.

FIMI can happen before or during significant periods of importance, such as during country elections, regional bloc elections (e.g., EU) or periods of political transition or contention. It is also important to note that the public should be aware of the domestic conditions within a country that could contribute towards the receptiveness of a population to FIMI attempts or narratives. It might be easier to create awareness on FIMI in populations or countries with high digital and media literacy levels. Countries or populations with lower or unequal levels of digital and media literacy might face greater challenges. Where the population has lower or unequal levels of digital or media literacy, fact-checkers play a crucial role in raising awareness of the potential risks of FIMI and highlighting the importance of digital literacy amongst the public.

Panel 2: Election Interference: Cases from Around the Globe

Disinformation and Democracy: Understanding Election Interference in Southeast Asia and Indonesia

Pieter Pandie, Researcher, Centre for Strategic and International Studies

- The rise of digital media and emerging technologies has transformed the electoral landscape in Southeast Asia, posing significant challenges to democratic integrity. Disinformation and foreign influence operations have become critical issues, particularly during elections. A study into disinformation and foreign influence operations on elections across Southeast Asia (with a detailed analysis of incidents between 2019 and 2024 in Indonesia, Australia, and Taiwan) revealed that there is a notable shift from text-based disinformation to video and audio formats - and from text- based platforms like X (formerly Twitter) to video and audio platforms such as TikTok and Instagram.
- This was true during Indonesia's 2024 election, making detection and verification significantly more difficult. This was compounded by low levels of information literacy among voters, with a national survey indicating that 96.7% of respondents had never attended digital literacy programs. Distrust in election integrity was also evident, further exacerbating the issue.
- The role of foreign influence operations was particularly prominent in the study, with digital platforms emerging as the primary channels for spreading disinformation. Such activities tend to intensify during periods of international conflict, as seen during the Russian invasion of Ukraine. Even in countries with robust policies against disinformation, such as Taiwan and Australia, the impact of these operations remains substantial.
- In Southeast Asia, addressing foreign influence operations is particularly challenging due to the topic's sensitive nature. Countries often hesitate to disclose cases of interference, fearing repercussions and undermining internal security. In Indonesia, for instance, although disinformation cases have been flagged, the domestic or foreign sources remain undisclosed, complicating mitigation efforts. Furthermore, influence operations are more effective in nations with weaker alliances or adversarial relationships compared to countries with strong international partnerships.
- During election periods, the sources of disinformation vary by country. In the Philippines' 2022 election, domestic actors were the primary sources of disinformation, whereas foreign sources were more prevalent in other Southeast Asian elections. AI-generated content further amplified the challenge, with deepfakes mimicking prominent political figures to mislead the public. These developments blurred the lines between authentic and fabricated information, complicating efforts to verify credibility.
- Low levels of independent information verification among the public further aggravate the issue of disinformation. In Indonesia, the habit of diligently cross-checking information is underdeveloped, with many individuals relying on superficial methods, such as quick searches on Google, for validation. This lack of verification skills and limited access to fact-checking resources leave the

population highly susceptible to manipulation. Public perception also diminishes the perceived severity of disinformation, as many view it as the work of individuals rather than orchestrated campaigns by larger organisations with ulterior motives. Such perceptions downplay the systemic nature of the problem, undermining the urgency of response measures. Disinformation also erodes trust in election-management bodies, reducing confidence in election integrity and weakening support for democratic systems.

- A multi-stakeholder approach is crucial to address these challenges. Governments, social media platforms, and civil society must collaborate to counter disinformation while safeguarding democratic freedoms. Enhanced cooperation between civil society platforms and domestic information agencies can improve resilience against disinformation campaigns. Emerging technologies, while contributing to the spread of disinformation, also offer opportunities for innovative countermeasures. Rapid advancements in AI necessitate the development of technological solutions that can effectively detect and neutralise AI-generated disinformation. However, these tools must be implemented alongside improved public education to enhance information literacy. Digital literacy programs should be prioritised to equip individuals with the skills to discern credible information from falsehoods.
- Governments must also strike a delicate balance between regulating the information landscape and upholding democratic freedoms. Overregulation risks stifling free speech, while under-regulation leaves societies vulnerable to manipulation. Ultimately, protecting electoral integrity in Southeast Asia requires a comprehensive and adaptive approach, leveraging both technological and educational initiatives to counter the evolving threats of disinformation and foreign influence operations.

Misinformation and AI - Lessons from the United States Election

Dr Samantha Bradshaw, Director, Center for Security, Innovation, and New Technology, American University

- Research into Gen AI's role in misinformation and disinformation during the 2024 United States election has illuminated significant challenges and potential opportunities in countering its misuse. Gen AI introduced new complexities in disseminating false information, impacting political discourse, public trust, and voter behaviour.
- There are four key challenges to note:
    - The increased persuasiveness of misinformation
    - The potential flooding of the information ecosystem
    - The personalisation of political messaging
    - The suppression of voter turnout through manipulative tactics
- The first challenge focuses on the persuasiveness of misinformation. Gen AI has enhanced the quality of disinformation by improving grammatical accuracy and linguistic sophistication, making it more convincing to specific target audiences. This is particularly evident in outreach to non-English-speaking communities, such as Spanish-speaking voters, where improved language proficiency in disinformation content has heightened its potential influence. However, despite

these advancements, academic studies and experiments suggest that human-generated disinformation remains more effective at persuading audiences than content produced by Gen AI. These findings temper some of the concerns about AI's immediate impact on the credibility of false information. There is also an argument that the creation of high-quality disinformation has long been achievable with pre-existing tools, raising questions about whether Gen AI significantly alters the landscape or merely accelerates trends already underway.

- The second challenge involves the sheer volume of misinformation that Gen AI can produce. The efficiency and low cost of the tools allow for the rapid generation of vast amounts of disinformation, potentially overwhelming the information ecosystem. Unlike traditional misinformation campaigns that rely on overt and detectable methods, such as copy-paste strategies, Gen AI enables more sophisticated approaches that can evade detection. This has raised concerns about the ability to identify coordinated behaviour and trace the origins of disinformation campaigns. However, the audience for most misinformation remains relatively small, often confined to highly partisan or conspiratorial groups. Hence, this may constrain the broader societal impact of increased misinformation volume, although its effects on these targeted groups can still be profound. The implications for local journalism and media credibility are particularly concerning, as the ability of journalists and fact-checkers to verify information and maintain public trust becomes increasingly difficult in this high-volume environment.
- The personalisation of political messaging represents the third major challenge. Gen AI allows for creating tailored disinformation campaigns, leveraging its ability to replicate behavioural patterns and mimic survey responses. This capability has been used to produce targeted political advertisements and messages that align with the preferences and biases of specific demographic groups. While personalisation can enhance engagement, it raises significant ethical and regulatory concerns about transparency and accountability in political campaigns. AI-generated content can obscure the origins of political messaging, making it difficult for voters to discern the motives behind the information they consume. Additionally, Gen AI can potentially suppress voter turnout through manipulative tactics. For example, malicious actors can use deepfake videos and audio clips to spread false information about candidates or voting procedures, misleading voters, or discouraging them from participating in the electoral process. Such tactics further undermine public confidence in democratic institutions and the integrity of elections.
- The election also highlighted broader, ongoing issues exacerbated by Gen AI, including media polarisation, erosion of trust in information sources, and the lack of transparency in AI-driven political advertising. These issues indicate the urgent need for long-term strategies to build societal resilience against disinformation. A key element of these strategies involves fostering greater empathy and understanding among diverse political and social groups, which can help reduce polarisation and improve the overall quality of public discourse. Tailored, localised responses to disinformation are also critical, as the effectiveness of countermeasures often depends on understanding the specific cultural, linguistic, and social contexts in which misinformation operates.

- The impact of Gen AI on the 2024 United States election underscores the need for a comprehensive approach to safeguarding democratic processes in the digital age. It is possible to create a more resilient information ecosystem by addressing the challenges posed by Gen AI while fostering collaboration, transparency and public awareness. This effort is essential for preserving the integrity of elections and ensuring that democracy continues to thrive in the face of rapid technological change.

Protecting Election Integrity against Disinformation: The European Experience

Professor Paolo Cesarini, Chair of Executive Board, European Digital Media Observatory (EDMO)

- The European Union (EU) has undertaken significant efforts to address disinformation and promote access to reliable information. These initiatives are grounded in recognising disinformation's broad implications, including its impact on security, democratic processes, and societal trust. The speaker emphasised the EU's regulatory framework, collaborative approaches, and the importance of fostering trust in media through long-term strategies.
- Disinformation poses security and societal challenges, undermining economic stability and eroding trust in democratic institutions. Therefore, it is imperative to establish reliable sources of information to uphold democratic rights. The evolution of information dissemination methods has transformed public perceptions and trust, necessitating innovative strategies to counter these effects.
- The EU has implemented various measures to combat disinformation, including regulatory actions and collaborative approaches. Transparency and accountability for online platforms have been prioritised through legal frameworks. Efforts to enhance the capabilities of public institutions in analysing and exposing disinformation campaigns are key components of this strategy. The EU adopted a whole-of-society approach involving diverse sectors such as media, civil society, academia, and e-commerce. This approach underscores the need to balance effectively countering disinformation and respecting democratic freedoms.
- Disinformation policies by the European Union were further coalesced in the aftermath of the downing of MH17 over eastern Ukraine. The MH17 case highlighted the need for coordinated responses to tackle disinformation. This led to the establishment of entities such as the European Digital Media Observatory and the involvement of the European Commission and European External Action Services in addressing disinformation. Regulatory measures have included the Digital Services Act (DSA), which mandates compliance among online platforms through provisions for penalties, sanctions, and independent audits. These measures aim to create a safer digital environment while upholding democratic principles.
- Regulating online platforms presents unique challenges, as excessive regulation risks undermining democratic values. Voluntary cooperation from technology companies is essential to ensure the effectiveness of regulatory measures. Multi-stakeholder engagement has been identified as a pathway for fostering

innovation and identifying manipulative information practices. Collaborative efforts between governments and technology companies have demonstrated success, such as during the COVID-19 pandemic, where coordinated campaigns reduced the impact of anti-vaccine disinformation. Social media platforms have also pledged to implement ethical standards to mitigate misinformation risks.

- The EU has sought to raise public authorities' capabilities by promoting intelligence-sharing among member states. Partnerships with technologically advanced countries like the United States are proposed as a means of creating automated systems to combat disinformation. These collaborative initiatives aim to strengthen defences against coordinated disinformation campaigns.
- Artificial intelligence (AI) introduces additional challenges in the fight against disinformation. AI-driven systems can amplify disinformation, complicating efforts to rebuild media trust. Legislation and cross-sectoral collaboration are essential to develop AI systems that promote media integrity. Professional expertise and targeted policy measures are required to ensure the long-term sustainability of efforts against AI-enabled disinformation.
- Policy directions for the future stress the importance of raising awareness and enhancing resilience against disinformation. The key components of this strategy are collaborative academic research and media literacy promotion. Funding support is critical to sustain initiatives and foster innovation in combating disinformation. Long-term strategy, regulatory measures, and partnerships across sectors will be essential to address the evolving challenges of disinformation and secure trust in information systems.
- By maintaining a balanced and inclusive approach, the EU aims to safeguard democratic values and ensure access to trustworthy information. These efforts reflect a commitment to building societal resilience and fostering an informed and engaged public in the digital age.

Key Points Noted from the Q&A Session

Issue: Challenges in Identifying Chinese Involvement in Elections.

Recent observations regarding pro-Chinese influence in elections reveal significant complexities in distinguishing between domestic and foreign disinformation campaigns. Much of the observed disinformation appears to originate from the political camps themselves, with no definitive evidence of foreign backing. This ambiguity complicates efforts to attribute responsibility and assess the extent of foreign involvement.

Should foreign actors be engaged in these activities, there are various challenges they would face:

- Bahasa Indonesia, a language that is not widely spoken outside Southeast Asia dominates Indonesia’s information landscape. This language barrier poses a significant obstacle to developing and disseminating persuasive disinformation by external actors.
- A successful campaign would require a nuanced understanding of Indonesia’s intricate political landscape, including its socio-political dynamics and cultural context, further raising the bar for foreign influence operations.

The current research has yet to pinpoint concrete examples of foreign involvement in political disinformation campaigns. Future studies focusing on the role of public relations firms and foreign actors in shaping information narratives may shed light on potential indicators of external influence. They could play a crucial role in strengthening awareness and improving detection of foreign interference, contributing to greater integrity and transparency in Indonesia's electoral processes.

Issue: Regulation and Legislation in Information Management.

A systematised legal framework is necessary to ensure consistent and effective oversight of information dissemination. Such legislation would help establish clear guidelines for managing digital platforms and combating disinformation. Internal processes and domestic sources play a key role in regulating the information environment. Governments can create tailored solutions that align with their specific political and cultural contexts by relying on local expertise and resources. This approach helps ensure that information management is effective and contextually relevant, reinforcing the need for strong, internal mechanisms in regulating the information space. Governments should also collaborate with platforms and push for increased funding and investments in "safety teams" to regulate the information space. Localised and context-dependent responses were identified as critical for effectively addressing information threats.

Issue: (In)Effectiveness of Fact-checking Labels.

Fact-checking labels, while intended to improve the accuracy of information, can sometimes reduce overall trust in various sources. This effect occurs because such labels may prompt scepticism, leading users to question not only the flagged content but other information from the same source. Additionally, the potential for confusion increases when content is falsely labelled, as users may disregard accurate information or misinterpret the intentions behind the labels. Social media platforms must invest more in developing advanced systems capable of detecting, preventing, and combating disinformation. These systems should go beyond simple labelling and focus on identifying patterns of misinformation, verifying content in real time, and providing users with transparent and reliable context to help them navigate the information landscape.

Issue: Broader Issues concerning Election Integrity and Disinformation.

Social welfare packages are sometimes used strategically to influence voter behaviour, blurring the lines between legitimate political campaigns and manipulative tactics. Addressing the root causes of disinformation campaigns and implementing measures that ensure the accuracy and transparency of electoral information is crucial to restoring election integrity. Strategic thinking is essential in tackling the long-term challenge of disinformation, requiring comprehensive efforts to safeguard the electoral process from ongoing threats. This involves immediate solutions and sustained efforts to strengthen public trust and resilience against future disinformation campaigns.

Panel 3: Emerging Technology (Gen AI) and Future Uncertainties

Gen AI in Strategic Communications: Use Cases and Models

Dr Doowan Lee, CSO, EdgeTheory; Social Media Task Force, Georgetown University

- Democracies are disadvantaged against malign actors in information operations who can engage in defensive and offensive activities. Democracies, on the other hand, can only rely on defence. However, there are exceptions. For instance, the United States Cyber Command is authorised to engage in both offensive and defensive information operations.
- Whether the impact of AI in communications is underhyped or overhyped remains to be seen. Synthetic content is not inherently bad, just as how organic content is not inherently trustworthy and good. However, AI can be detrimental to democracies. Examples of this include the use of deepfakes in the Slovakian elections in 2023 (built by AI tool ElevenLabs), Bangladeshi elections in January 2024 (built by AI tool HeyGen) and others. For instance, the deepfake featured an international news site criticising Bangladesh's policy on Gaza and was released a day or two before the polls opened. The content was viewed millions of times by the time the polls were open.
- The Green Cicada Network employed in Australia is a good example of how powerful AI-assisted disinformation campaigns can be. Within 3 months, the 3,054 fake accounts that are part of the network became the most engaged with accounts on X. The network was detected because these accounts were activating the safety trigger on X. The network only took one programmer, who was an AI researcher working at a lab in Qing Hai University, to build and run, showing how AI can be used in amplifying the scale of disinformation operations. AI can also transform the use of troll farms in mass disinformation operations, with AI replacing this workforce.
- Another issue with AI is the difficulty surrounding its regulation/governance. It is difficult to put guardrails around AI, and certain troubling indicators point to the fact that we are not prepared to consider what AI can do in the next couple of years and the implications of that. Since November 2022, 1 trillion USD has gone into the Gen AI industry, encouraging its dynamic growth and predicting for developments to come much faster in the years to come.
- However, there are attempts in the United States to put guardrails around larger LLMs. The Coalition for Content Provenance and Authenticity (C2PA), and Google's Sync ID (a watermarking protocol for synthetic content) are examples. Despite attempts, many unsanctioned LLMs are traded in the dark web (e.g., WolfGPT, XXXGPT, FraudGPT etc.), and there are limited mechanisms to deal with this.
- The use of AI models in information operations and communications will only accelerate, and relying on the detection and analysis of adversarial use cases alone is unlikely to favour the defender. There is a growing awareness that defence alone is not enough. There is a need to re-examine the balance between offence and defence aspects of communications and how we can become more

proactive with strategic communications amidst the competition for attention. AI can be a highly effective tool to scale strategic communications, and such solutions should be considered because plain defence would be futile in the age of Gen AI. The use of AI infrastructures can tip the balance between aggressor and defender.

Examining the Effects of LLM-powered search on Political Knowledge and Engagement

Dr Kokil Jaidka, Assistant Professor, Communications and New Media, National University of Singapore

- AI has been used to intentionally create problems for democratic processes, behaviours, perceptions, and others. There are also instances where the unintentional use of AI contributed to harms against democracy, especially with the growth of AI use. Search engines are increasingly forcing users to interact with chatbots, and the study presented explores what this means for users looking for news.
- Across the West, there are large swathes of territories with no/inadequate local news coverage. This has created an issue with representation regarding chatbot results, even in the United States, where search engines are based. News curated by search engines is thus not representative of and/or ignores local contexts. This has critical implications, particularly during evolving crisis events like protests and elections. For instance, voters deciding on local representatives based on nation-wide information could present an issue for democracy.
- There are research gaps in the quality of political news presented by chatbots, user perception of news consumed via chatbots, and research on chatbots beyond English.
- Results from the study revealed that out of the queries posed on ChatGPT and Google News on Singapore news, Google News results had 2/23 relevant news while ChatGPT returned 0/23 relevant news. In Indonesia, 66/184 results were returned from queries to Google News, and 31/184 results were returned from ChatGPT. The numbers from the United States revealed 147/2249 results from Google News and 279/22346 from ChatGPT for news about local politics.
- The results suggest that relying on chatbots for local news is not ideal. However, users interacting with English-language chatbots view chatbots as less biased, more credible, more relevant, and more understandable. They also spend more time on these chatbot-based search services. Despite questionable differences in quality, people do prefer chatbot-enabled news consumption.
- However, while interacting with chatbots improve knowledge for English-speakers, for Spanish-speakers interacting with Spanish-speaking chatbots, for example, it reduced their ability to detect misinformation. As such, not only is the data less representative towards non-English languages, but the information presented by non-English chatbots may also be actively harmful. The limitations of existing Application Programming Interfaces (API) in fetching diverse news leads to LLMs' limited and outdated political information retrieval, particularly in the case of non-English language news. Currently, the presented study will check chat transcripts for bias, to examine if trust in chatbot-powered answers

has facilitated confirmation bias. Users' inability to audit the backend processes of LLM has also led to trust in chatbot-powered answers.

- Increase in audits, transparency and multilingual evaluations are necessary to ensure equitable access to credible information. There is a need for more exploration across contexts to see how LLM search can improve information quality. There is a need for more interventions to allow users to understand the information they are consuming, especially when dominant narratives of AI harm focus on more "scary" AI harms.

Gen AI: Opportunities and Challenges for Misinformation and Safety

Dr Priyanka Bhalla, Head of Safety Policy, APAC, Meta

- Meta seeks to fight misinformation by (a) removing content, (b) reducing the distribution or virality of content or (c) informing users on Meta's actions on this issue.
- On removal of content, Meta cannot remove all the misinformation because of the following reasons:
    - As an American company, it does not see itself as an arbitrator of truth,
    - There are differences in reception about what is true and false, and Meta should not be the decision maker on such content,
    - Given misinformation is often shared unknowingly, Meta does not want to penalise a person for an act they have done unknowingly.
- Meta's global community guidelines guide its actions on misinformation. The company aims to balance between safety and respect in online spaces with freedom of speech. Misinformation responses on elections have mostly targeted misinformation that might cause "violent imminent harm" or that aimed at "voter suppression".
- Reduction action seeks to curb the distribution or virality of false or altered content, which may involve the reduction of the distribution of pages or domains that have repeatedly disseminated false content. Inform action, on the other hand, involves placing a fact-check label on content to inform users that the content might have been manipulated. If the content is from a repeat offender, Meta provides a notification to inform users. Meta works with a global network of 100 certified fact-checkers that cover 60 different languages and 37 countries in the Asia Pacific (APAC).[1]
- There was an overwhelming focus on deepfakes during elections. However, there were not many deepfakes in circulation. Besides, there has been advancement in providing disclosures and labelling AI-generated content. Meta, in partnership with other industry peers, identifies different types of content and puts visible and invisible watermarks on them. On election-related content such as digitally created or altered ads, Meta encourages the disclosure of AI-generated content.

[1] Please note that the speaker delivered this presentation in November 2024. This piece of information shared in the presentation was accurate at the time of the conference (November 2024). Meta announced it would halt its fact-checking programme starting with the United States in January 2025.

- There were also cases of positive uses of deepfakes during elections. For instance, former Prime Minister of Pakistan, Imran Khan, managed to reach different audiences during the election period via AI-generated videos that were labelled as AI-generated.

Key Points Noted from the Q&A Session

Issue: The offense and defence balance in information manipulation activities will change with technological advancement.

Multimodality in AI, such as conversions from image to audio or from text to audio, has been central in 2024, while chatbots were under the spotlight in 2023. Switching between different modalities of content generation is becoming easier. In the upcoming 2024-2025 phase, Agentic AI[2] will take centre stage and with this, there will be a shift from feeding AI models with prompts to generate content to assigning them a task. For instance, an actor will be able to ask the AI model to develop a disinformation campaign with an automated workflow. The model will then create content and fake social media accounts, and disseminate the narratives. Researchers will likely struggle to keep up with this development and advancements will likely nudge the offence-defence balance. In this context, being proactive will be important. An example is the difference in Ukraine's approaches to information sharing in 2014 and 2022 – Ukraine has been proactively sharing content since the 2022 Russian invasion.

Issue: The dependence on a limited number of dominant AI models can have consequences.

The Computers and Society Journal had an article where the authors asked the six most popular large language models (LLMs) who caused the war in Ukraine and received diverse responses from different models. However, when asked the same question in Russian, the variation decreased, suggesting that the models are language and context sensitive. It is essential to move from depending on a limited number of American models to building regionally tuned, applied foundation models, especially to avoid any representation diversity-related problems. Specific to Meta on this topic, the company does red teaming to ensure search results do not lead to bad content and that searches prompted in Meta products offer a variety of outcomes. Meta works with different organisations to combat bad content.

Issue: Whether Meta has measures in place to encourage self-disclosure of AI-generated content and to combat scams.

If a third party shares content without self-disclosing that it is AI-generated, especially for advertisements, Meta will take down the content. If the entity is a part of the Coalition for Content Provenance and Authenticity (C2PA), they have agreed to follow specific standards and ethics. While some argue that open source LLMs are less safe, Meta found the opposite. On scams, Meta has fraud and deception policies and

[2] "Agentic AI uses sophisticated reasoning and iterative planning to autonomously solve complex, multi-step problems". Please see Erik Pounds, "What Is Agentic AI?", *NVIDIA*, 22 October 2024, https://blogs.nvidia.com/blog/what-is-agentic-ai/.

strategies to tackle scams. It is also currently pilot testing a new face recognition technology targeting celebrity-based scams to ensure it is the real person captured in the advertisement.

Issue: Gen AI will continue to develop at fast speed, making us reconsider the possible guardrails.

There are some existing guardrails on Gen AI development, and their presence is significant. However, the rapid advancement of Gen AI will not slow down, especially when the field attracts significant investment and under the new administration in the United States. Besides, there is a geopolitical angle to the fast development of Gen AI, with the ensuing competition of the United States and China in the field. China is making rapid advances in Gen AI. When the United States delegation visited China about two years ago, the country did not have any foundational models. However, China has the capacity to invest in strategic industries and it achieved 4 Large Language Models in less than two years. Guardrails are necessary against this fast pace of development and when people are losing their lives to inaccuracies and problematic language models. However, there is a lack of incentives to scale trust and safety and tech policy within the current geopolitical context in which Gen AI develops. The solution lies in countries developing their language models and safe use cases and solutions. While the European Union may be more suited than others to offer top-down guardrails or policy interventions, it is not well positioned to build its models due to a few factors. These include: (a) high energy prices, (b) Insufficient human capital, (c) lack of institutional support.

Panel 4: Case Studies: Exploring Platforms, Targets, Tactics and Countermeasures

FIMI in Malaysia: Understanding risks and ways forward

Harris Zainul, Deputy Director (Research), ISIS Malaysia

- Malaysia is a highly networked nation with high internet connectivity and social media usage. There is a need for better localisation of the FIMI concept as the concept is based on risk profiles of other countries, which may be non-applicable for Malaysia. The country's FIMI risk profile may differ from other countries due to Malaysia's neutral foreign policy stance.
- There is also a need to distinguish foreign information manipulation and interference with legitimate foreign activity, such as international donors supporting local human rights groups, NGOs and foreign media outlets reporting on Malaysia. Malaysia should increase its capacity to detect and attribute FIMI operations. The country should do this deliberately to protect the growing democratic space.
- While the origins of influence operations and misinformation may be foreign, domestic actors can equally be responsible for undertaking information operations or engaging in coordinated, inauthentic behaviour against the population. This may include using trolls and influencers on social media and incorporating traditional media outlets.
- The risks of FIMI to Malaysia are:
    - Elections: While Malaysia's democracy has become more competitive than before, having moved from one-party rule for 61 years to four different governments in the past six years, there is no present instance of FIMI based on observations from the 2022 elections.
    - Societal relations: Malaysia has a multiracial and multireligious population, with inter-group relations more of a compromise than having a united national identity. Sensitive topics are generally avoided in public discourse. This could present opportunities for mistrust among racial groups to go unaddressed. Adversaries may exploit this lack of resilience.
    - Geopolitics: Malaysia has outstanding territorial disputes that can be potentially exploited for FIMI operations. Events and conflicts further away may also result in FIMI operations, as seen in the aftermath of the downing of Malaysia Airlines' MH17. FIMI operations and disinformation could potentially skew public opinion on these issues, either to confuse, deflect blame, or aggravate. Nonetheless, FIMI operations are unlikely to change minds, but they may solidify existing beliefs.
- The FIMI countermeasures available to Malaysia include:
    - Laws and regulations: Existing laws govern parts of FIMI, including improper use of network facilities and the making of statements conducing to public mischief.
    - Platform governance: Platforms with more than eight million users in Malaysia must be licensed to operate from January 2025. This requirement was instituted after the Malaysian government assessed

that the platforms have not been diligent enough in addressing online harms. These licensing conditions include conducting "*regular assessments of systemic risks*," submitting half-yearly online safety reports detailing measures taken to enhance online safety and having a dedicated local content moderator team with adequate support and training.

- There are concerns over these countermeasures, such as the standards of moderation by platforms, government overreach in balancing between online safety and free speech considerations, and the level of compliance by platforms.
- It is more important to improve the population's resilience to combat FIMI. This could include more sophisticated media literacy and public awareness campaigns with more targeted programmes aimed at different demographics and communities, more organised and professional fact-checking to streamline resources, and cross-post fact-checks to reach wider audiences as well as pre-bunking to inoculate Malaysians against mis- and disinformation. However, these are hard to scale, and it is difficult to target the most vulnerable groups to FIMI. Financing these measures through non-government organisations is also tricky.

Gen AI and Foreign Disinformation in the 2024 U.S. Election

McKenzie Sadeghi, AI and Foreign Influence Editor, NewsGuard

- AI has exacerbated misinformation and changed the journalism landscape with AI tools that readily spread disinformation. State sponsored information operations have also used AI to increase the scale and persuasiveness of their messaging. AI impersonating credible media has also been used by malicious actors to exploit and undermine trust.
- The AI-generated misinformation landscape is monitored using different means, such as red-teaming current AI tools to assess potential weaponisation, tracking real-time proliferation of AI-generated news sites using social media analytics tools, and leveraging data and open-source reporting techniques to spot AI-generated photos, video, and audio.
- AI-generated news sites are defined by the following characteristics:
    - The presence of clear evidence that AI produces a substantial portion of the site's content.
    - The presence of substantial evidence content is published without significant human oversight.
    - The site presented in a way that an average reader could assume that human writers or journalists produce its content, and
    - The site does not disclose that AI produces its content.
- There has been evidence that malicious actors are already offering AI-generated sites posing as news outlets, such as reposting misinformation on a Russian network of sites that purportedly look like local news media sites in the United States. Gen AI has been used to populate this fake news network with news articles and has dramatically reduced the operational workforce needed to operate the whole network of disinformation sites and eliminated the need for in-country costs.

- Experts have also observed an AI disinformation cycle where Gen AI models mimic disinformation claims and cite the Russian-created fake news sites as authoritative sources. This exacerbates the effect of deepfakes, where "exposés" are uploaded onto YouTube, and AI-generated articles are based on these contents, often accompanied by AI-generated images. Chatbots then repeat this AI-generated narrative from the disinformation network, leading to the potential for more deepfakes. Iran has also been noted to allegedly utilise similar tactics.
- Tackling disinformation will be more difficult with the progress of Gen AI. Deepfakes are increasingly becoming more sophisticated and convincing, with greater detail and reality. Foreign actors can use these AI models as *casus belli* (i.e., an act or situation that provokes or justifies a war) and amplify false narratives. AI has also exacerbated the "pink slime" problem in local news outlets and the number of these outlets has surpassed the number of authentic local daily newspapers in the United States.

Cases of FIMI in Vietnam: Exploring Different Platforms, Targets, and Tactics

Dr Viet Tho Le, Deputy Dean, School of Media and Applied Arts, University of Management and Technology, Ho Chi Minh City, Vietnam

- The rapid adoption of social media such as Facebook, TikTok, and YouTube in Vietnam has exposed the population to the risks of foreign interference and misinformation. Foreign information manipulation and interference threaten public trust, manipulate opinion, and impact social and policy stability.
- Malicious actors use political narratives, COVID-19 misinformation, and sensationalised news to disseminate FIMI on social media platforms, and it has been amplified through viral content. Each social media platform has specific tactics and challenges regarding FIMI. Facebook's algorithm promotes high-engagement content that could involve misinformation and can be used in coordinated inauthentic behaviour campaigns. TikTok's viral, short-form videos create an environment for misinformation to spread, and curated content to a user's history can fuel echo chambers. YouTube also suggests recommendations to users, creating a "rabbit hole" effect while monetising content and incentivising the sensationalisation of content.
- Vietnam currently has some regulatory instruments to combat FIMI. These include fines for false information posting, government requests to social media platforms for content removal, requirements for local data storage, and the regulation of influencers and high-profile account holders. New data regulation and account verification requirements will kick in on 25 December 2024, mandating social media accounts to be linked with verified phone numbers or a personal identification number for posting to eliminate anonymous FIMI dissemination. These measures seek to increase the accountability of users but raise concerns over privacy and free speech.
- Due to the adaptive and amorphous nature of FIMI, Vietnam needs a proactive approach to tackle the issues. This will require continuous policy updates and should keep in mind balancing regulation with freedom of expression in a secure digital environment. This balance can be attained by Vietnam improving its

digital literacy, having a collaborative regulation environment, ensuring clear transparency and accountability guidelines, and strengthening data privacy protections.

Key Points Noted from the Q&A Session

Issue: FIMI is a regional concern across ASEAN and countries should collaborate to counter cross-border disinformation.

To an extent, countries can aim to harmonise rules, standards, and definitions regarding FIMI, misinformation, and disinformation. Some headway on harmonisation was made during Indonesia's chairmanship of ASEAN in 2023. However, despite parallels between certain ASEAN countries (e.g., in societal relations and demographic makeup), differences in economic relations and foreign policy dictate different risk profiles, which should also be appreciated. While countries may share best practices and research concepts, these must be suitably localised to fit each country's unique context and circumstances. Most notably, ASEAN countries can consider similar harmonisation efforts between European Union member states in strategies to tackle FIMI.

Issue: The usage of troll farms and cyber troops to influence political discourse in Malaysia and if new legislation to address FIMI is required.

Troll farms and cyber troops are now part of most Malaysian political party campaign strategies. Notably, the exchange of funds for digital labour purposes (in this case, political campaigning) is now mainstream and no longer stigmatised. Transparency regarding this is an increasingly significant concern, and self-disclosure is unlikely to be effective. Introducing new legislation is also unlikely to be effective, as existing legislation can already be applied if necessary. Moreover, it may not be possible to enforce legislation against threat actors located beyond Malaysia's territorial jurisdiction. As such, ensuring transparency and sufficient oversight (especially on social media platforms, where FIMI activities commonly occur) should instead be emphasised.

Issue: Policies by social media companies are crucial in tackling the emergence of new challenges from the rise of AI.

Due to the sophistication of AI and chatbot technology, it is increasingly difficult to detect AI 'pink slime' outlets. The large increase in the number of AI-generated websites poses a further challenge in combating AI disinformation. While many of these sites may not necessarily publish disinformation, the large amount of low-quality content propagated by these sites is used to attract web traffic. Government regulation may not always be relevant or effective. Instead, actors could be disincentivised from attempting to generate revenue from these sites or social media platforms. This would, in turn, require more transparency and oversight regarding social media activity and greater effective cooperation between states and social media platforms.

Issue: The effectiveness – and unintended consequences – of Vietnam's countermeasures against false information.

In addition to implementing cybersecurity and data laws, Vietnam has appointed a digital military force with an active online presence and officials to meet with leaders of prominent social media platforms, who direct them to remove certain forms of content. Domestic advertisers are also asked not to pay money to advertise on specific platforms. However, the government's approach to fake news and false information may lack clear definitions. For example, the arbitrary framing of certain voices as 'hostile' or foreign' can sometimes make regulation incoherent and ineffective. Furthermore, digital literacy in Vietnam remains weak overall.

Issue: Improving the state of public trust and credibility in journalism.

Media organisations can aim to provide readers with information about the source of news they are reading, perform due diligence regarding transparency, and aim to divert advertising revenue back to traditional, reliable, and reputable news outlets. Information-sharing initiatives between countries may also allow for closer monitoring of disinformation networks. In ASEAN, this may involve establishing an ASEAN regional centre for combating fake news and countries jointly developing an AI detection software.

About the Centre of Excellence for National Security (CENS)

The Centre of Excellence for National Security (CENS) is a research centre that studies, publishes, and speaks publicly on national security areas including cybersecurity, cyber conflict, disinformation, online harms, hybrid threats, foreign interference, economic interference, social resilience, radicalisation, and the impact of technology (including emerging technologies like artificial intelligence) on them, especially from the perspective of small states and non-western regions such as Singapore and Southeast Asia.

For more details, please visit www.rsis.edu.sg and www.rsis.edu.sg/cens.

Join us at our social media channels at www.rsis.edu.sg/rsis-social-media-channels or scan the QR code.