April 14, 2022

IRSEM
INSTITUT DE RECHERCHE STRATÉGIQUE
DE L'ÉCOLE MILITAIRE

RESEARCH PAPER – No. 126

TRENDS IN LEGISLATION AGAINST FALSE AND HARMFUL INFORMATION IN THE ASIA-PACIFIC

Celine THAM
Visiting research fellow at IRSEM

ABSTRACT

While "fake news" as a phenomenon is not new, today's digital media age has made the need to address it considerably more urgent. Its political impact, potential to compromise the integrity of electoral processes, and ability to cause real world harm have driven governments across the globe to take notice. The trend towards legislation as a countermeasure is unmistakable, with many new pieces of regulation targeting the creation, distribution and manipulation of false and harmful information being enacted in the last four years, and many more still being drafted and considered. This paper maps and compares the regulatory frameworks for addressing false and/or harmful information in five economies in the Asia-Pacific – Australia, India, New Zealand, Singapore, and Taiwan. Its aim is to demonstrate the diversity of regulatory strategies which have been implemented or are under consideration, and in doing so, act as a discussion starter on governance of the digital space, where the circulation of ideas could better inform the fight against false and harmful information, which spreads not just within but across national borders.

CONTENT

MINISTÈRE DES ARMÉES
Liberté
Égalité
Fraternité

DISCLAIMER: One of IRSEM's missions is to contribute to public debate on issues relating to defence and security. The views expressed in IRSEM's publications are the authors' alone and are in no way representative of an official Ministry for Armed Forces stance.

INTRODUCTION

"Fake news" as a phenomenon is not new. However, the digital media age of today enables a wider range of potential perpetrators, as well as facilitates a much broader distribution of information to a global audience, and has made the need to address it considerably more urgent. In addition to being inscribed in the public consciousness, the issue became the subject of heightened focus of governments and a political issue in the wake of the 2016 US Presidential Elections, as its potential to compromise the integrity of electoral processes and its political impact was widely evidenced.[1] In step with governments across the globe placing the need to counter the issue high on their agenda, many legislators have taken to the challenge of regulating false and/or harmful information, which poses a threat to public interests. There has been an emergence of new laws across the globe, with at least 28 countries having passed legislation related to false news as of March 2020, either by passing amendments to existing regulations or enacting new legislation altogether.[2] Since then, many more have criminalized the act of spreading false news as a response to the COVID-19 infodemic, with the Atlantic Council's DFRLab accounting for at least 24 countries having done so.[3] The Asia-Pacific region is no exception.

This paper maps and compares the regulatory frameworks for addressing false and/or harmful information in five economies in the Asia-Pacific – Australia, India, New Zealand, Singapore, and Taiwan. While not claiming to be an exhaustive stock take of the region's attempts to counter false and harmful information, it captures the varied manner in which the problem has been approached, and covers the range of regulatory tools adopted which cut across different Internet policy areas and actors, with a focus on content which hitherto would have been considered lawful. Its aim is to demonstrate the diversity of regulatory strategies which have been implemented or are under consideration, and in doing so, act as a discussion starter on governance of the digital space, where the circulation of ideas could better inform the fight against false and harmful information, which spreads not just within but across national borders.

HOW THE PROBLEM IS DEFINED

Although the phenomenon has been the subject of considerable scientific research and political debate in recent years, there is no universally accepted lexicon to describe it. Ironically, the plethora of terminology – fake news, propaganda, information warfare, hoaxes, mis-/dis-/mal-information and information manipulation, amongst others – and the difficulty in formulating precise definitions of the subject has been well-documented

[1] Jean-Baptiste Jeangène Vilmer, "Information Defense: Policy measures taken against foreign information manipulation", Digital Forensic Research Lab, Atlantic Council, July 2021, p. 2.

[2] Kalina Bontcheva et al., "Balancing Act: Countering Digital Disinformation while respecting Freedom of Expression", Broadband Commission research report on 'Freedom of Expression and Addressing Disinformation on the Internet', September 2020, p. 108.

[3] Jacqueline Malaret and John Chrobak, "The criminalization of COVID-19 clicks and conspiracies", Digital Forensic Research Lab, Atlantic Council, May 13, 2020.

by scholars. While some consensus regarding the nomenclature appears to be emerging, most notably the rejection of the term "fake news", in part because it has been increasingly instrumentalized by politicians to describe news coverage disadvantageous to themselves, there is less agreement regarding the typology of the problem.[4] Questions concerning the dichotomy between truth and falsehood, malicious intent as a defining criterion, as well as the relationship between problematic behaviours and the veracity of content – against an acknowledgement of the inherent fluidity of these concepts, where the timing at which the same content is spread can have varying consequences – have not yet been resolved.[5]

Unsurprisingly, this ambiguity is similarly observed across the places surveyed in this paper and applies even at the most basic level, where a multiplicity of terms appears to describe the same manifestation, viz. information enabled by digital media and which fulfil three criteria: false or misleading, spread with malicious intent, and has the potential consequence of causing harm to a person, a group, organisation, society or country. While Australia, New Zealand and Taiwan's preferred term is "disinformation", Singapore and India have opted for "deliberate online falsehoods" and "fake news" respectively. If such information is known to be of foreign provenance, Singapore labels it a subset of a "hostile information campaign", while Taiwan qualifies it as disinformation "at the instruction or commission of or with financial support from hostile external forces". In the other three places, the nomenclature is either less defined or coupled with the related issue of foreign interference.

Mirroring the academic debate, what quickly complicates the difference in terminology are the additional nuances concerning what constitutes harmful information which should be actively managed. The Australian Communications and Media Authority (ACMA) has already called into question the adequacy of the traditional intent-based definition of disinformation in encapsulating the breadth of the issue, because ordinary users can spread harmful and misleading content unwittingly.[6] The line between malign actors and benign users is further blurred when malicious disinformation campaigns seek to co-opt such ordinary users to inadvertently promote particular narratives.[7] As such, misleading information shared even without the intent to cause harm, what ACMA terms as "misinformation", can still cause significant harm. Singapore and Australia have also challenged the notion that harmful information must also be false. The former recognises the exploitation of sensitive issues to polarise views and turn people against one another as a feature of hostile information campaigns.[8] Similarly, the latter acknowledges that accurate information

[4] Alexandre Alaphilippe et al., "Automated tackling of disinformation", March 2019, p. 5-7.

[5] Jean-Baptiste Jeangène Vilmer et al., *Information Manipulation: A Challenge for Our Democracies*, report by the Policy Planning Staff (CAPS) of the Ministry for Europe and Foreign Affairs and the Institute for Strategic Research (IRSEM) of the Ministry for the Armed Forces, Paris, August 2018, p. 18-22; Andrea Carson and Liam Fallon, "Fighting Fake News: A Study of Online Misinformation Regulation in the Asia Pacific", La Trobe University, January 2021, p. 47-50.

[6] Australian Communications and Media Authority (ACMA), *Speech by Nerida O'Loughlin, ACMA Chair, RIGA Stratcom dialogue May 2021*, May 10, 2021.

[7] ACMA, *Misinformation and news quality on digital platforms in Australia: A position paper to guide code development*, June 26, 2020, p. 9-10.

[8] Ministry of Home Affairs (MHA), *Committee of Supply Debate 2021 on "Keeping Singapore Safe in the Evolving Security Environment" – Speech by Mrs Josephine Teo, Minister for Manpower and Second Minister for Home Affairs*, March 1, 2021.

inappropriately spread by bad-faith actors (termed "malinformation") can be detrimental, particularly to the operation of democratic processes.[9] It is worth noting that ACMA ultimately categorises malinformation as a feature of coordinated disinformation campaigns, rather than an independent phenomenon, which speaks to the complexity and the accompanying difficulty in formulating strict definitions of – and by extension, interventions to – the problem.

This difficulty in disentangling the problem is further compounded by other categories of harmful content, similarly exacerbated by the proliferation of digital and social media platforms. In view of characteristics such as being difficult to objectively discern and tending to produce diffused societal impacts rather than narrow personal harms, it has been argued that there has been an impulse to regulate the identification and enforcement of different categories of harmful information under a single, comprehensive, regulatory framework.[10] This is well represented by New Zealand's initiation of a "broad, harms minimization-focused" media content regulatory review to update its regulatory system to respond to "many digital media types".[11] Led by the Department of Internal Affairs (DIA), the review aims to minimise the likelihood of New Zealanders unintentionally coming across "harmful content", a catch-all term which comprises eight vastly different categories of legal and illegal content ranging from adult content which children can access, child exploitation material, disclosure of personal information that threatens someone's privacy or promotes self-harm, unwanted digital communication, mis/disinformation, racism and other discriminatory content, as well as hate speech.[12] The Indian Ministry of Electronics and Information Technology (MeitY) has adopted a similar approach, having rolled out a single set of Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, which imposes obligations on digital intermediaries to address the spectrum of "disturbing developments" on social media platforms.[13] This tendency could also be explained by the fact that interventions for offences made possible by the Internet and digital technology are often similar. For example, whether it be cyber bullying, disinformation, or the spread of violent extremist content, taking down the content or disabling public access to it is a common solution to arrest its propagation.

Regardless of the definitions and characterization of the problem, there appears to be unanimous agreement that the integrity of the information space is being challenged and needs to be safeguarded, and that digital platforms cannot be left to self-regulate.

[9] ACMA, "Position paper", p. 10.

[10] Jason S. Pielemeier, "Disentangling Disinformation: What Makes Regulating Disinformation So Difficult?", *Utah Law Review* 2020, no. 4, p. 921.

[11] Department of Internal Affairs (DIA), *Proactive release of Cabinet material about the initiation of the Content Regulatory System Review*, July 2, 2021.

[12] Ibid.

[13] Press Information Bureau (PIB), *Government notifies Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021*, February 25, 2021.

AN ACCOUNT OF THE LEGAL LANDSCAPE

All five places examined in this paper have put in place regulatory interventions to govern the online domain, revolving around similar sets of harm, albeit with varying orders of priority. While this paper is primarily focused on legislation directed at false and harmful information, spread intentionally or otherwise, the ensuing account of the respective legal frameworks also includes those targeting hate speech, cyber bullying, extremist content, and foreign interference, to provide a broad overview of the different schools of jurisprudence and how they evolved to deal with the threat of false and harmful information. It is in no way intended to be an exhaustive list of legislative levers which could be or have been invoked against false and harmful information, particularly as the landscape will likely change over the coming months. The succeeding section also examines the discourse and legislative intent put forth by the respective authorities and their specific contexts, to contextualise the legal framework and strategies that have been adopted.

Australia

From a regulatory standpoint, the Australian Government has thus far broadly framed and approached the issue of harmful and false news enabled by digital media in two ways.

First, within the broader context of foreign interference and its impact on Australia's sovereignty, values and national interests. When then-Prime Minister Malcolm Turnbull first introduced new legislation to counter foreign interference in December 2017, he quoted the Australian Security Intelligence Organisation (ASIO) in stating that the threat was "unprecedented", and outlined cases of "covert, coercive or corrupting" behaviour, a framework now known as the "three C's", to exert improper influence over governments and political landscapes being observed globally, noting that "questions of foreign interference are not all about China".[14] He dedicated a significant portion of his speech to how such behaviour sought to "manufactur(e) public opinion to hijack political discourse and tilt the decision-making landscape to their advantage", with the internet being leveraged as a "turbocharged" instrument of division and to democratize disinformation.[15]

A new *Foreign Influence Transparency Scheme* was thus introduced and passed the Parliament in June 2018 to impose registration requirements on persons undertaking activities – including communications activities – on behalf of foreign principals or for political or governmental influence, with the objective of ensuring transparency behind these activities, allowing the public and policymakers to assess any underlying agenda. At the same time, *the National Security Amendment (Espionage and Foreign Interference) Act 2018* criminalised acts of foreign interference which could influence a political or governmental process, influence the exercise of an Australian democratic or political right or duty, support

[14] Malcolm Turnbull, *Speech introducing the National Security Legislation Amendment (Espionage and Foreign Interference) Bill 2017*, December 7, 2017.

[15] Ibid.

intelligence activities, or prejudice Australia's national security. Turnbull said that these legislative tools, coupled with equipping the government with the ability to enforce them, will allow Australia to deal with the risks "surgically" and send an "unmistakable signal" of deterrence.[16] It has been suggested that this focus on countering political disinformation is unsurprising, considering the vagaries of Australian politics, where mid-term leadership changes have become the norm, offering ample opportunities for foreign interference.[17] While the federal government did not name a specific country in the passing of the 2018 laws, and have sought to underscore that it adopts a country-agnostic approach in countering foreign interference, they were introduced around the same time as the resignation of Labor Party senator Sam Dastyari for his alleged close ties to a prominent Chinese businessman and political donor, and pro-China stance on the tensions in the South China Sea, which precipitated the broader public debate about foreign meddling in Australian politics.[18]

Reflective of the two laws being more comprehensive in the range of foreign interference activities beyond that of false and harmful information that it seeks to ensnare, the only prosecution to date is under *the National Security Amendment (Espionage and Foreign Interference) Act 2018*, and is likely a case of espionage which has yet to fully play out in the courts.[19] There are several other cases of alleged foreign interference offences reportedly under investigation, with the case involving an act of covert influence towards and lobbying of an Australian politician being the first to grab public attention in 2020.[20]

Australia's ongoing efforts to manage false and harmful information have secondly been scoped to content presented as news and journalism. This is guided by the Australian Competition and Consumer Commission (ACCC)'s concern that a large proportion of the population risks being exposed to poor quality news, given that the Internet has become their primary source of news, and that the ability to recognize high-quality news is essential for a well-functioning democracy.[21] This was a key finding which arose from the ACCC's 2019 inquiry into the effect that digital search engines, social media platforms and other digital content aggregation platforms have on competition in media and advertising services markets, as directed by the Turnbull administration.

Within this context, the ACCC recommended the implementation of an industry-developed code of conduct to govern the handling of complaints about disinformation that meet a "serious public detriment" threshold, and that an independent regulator such as the ACMA should supervise and enforce the code, as well as other voluntary initiatives

[16] Ibid.

[17] Kanchan Kaur et al., "Information Disorder in Asia and the Pacific: Overview of Misinformation Ecosystem in Australia, India, Indonesia, Japan, the Philippines, Singapore, South Korea, Taiwan, and Vietnam", the Journalism & Media Studies Centre, University of Hong Kong, March 3, 2019, p. 58.

[18] Ben Westcott and Serenitie Wang, "Australian senator resigns over allegations of Chinese influence", *CNN*, December 12, 2017.

[19] Erin Pearson, "Factory of first person charged under foreign interference laws bugged as part of ASIO probe", *The Age*, April 6, 2021.

[20] Sean Rubinsztein-Dunlop and Echo Hui, "Australian police accessed Chinese diplomats' emails and messages as part of foreign political interference investigation", *ABC News*, Sep 15, 2020; Rob McGuirk, "Australian lawmaker says he isn't a suspect in Chinese probe", *AP News*, June 29, 2020.

[21] Australian Competition & Consumer Commission (ACCC), *Digital Platform Inquiry Final Report*, July 26, 2019, p. 342-358.

which enable users to identify the quality and source of news content.[22] ACMA subsequently issued a position paper to guide the development of the code, asking that it also address harmful misinformation in view of the 2019-2020 bushfire season and the COVID-19 pandemic, which showcased the significant harm of false information shared even without malicious intent.[23] It also advocated a "graduated and proportionate" approach, where various measures could be implemented based on the assessed likelihood and severity of harm, as well as taking into consideration the unique characteristics of the various platforms.[24] In February 2021, platform providers including Twitter, Google, Facebook, Microsoft, Redbubble, TikTok, Adobe and Apple adopted the *Australian Code of Practice on Disinformation and Misinformation ("Code")* and issued the first set of transparency reports in May 2021.[25] ACMA has submitted to the Government its assessment on the code and the platforms' initial compliance, and the latter has provided preliminary indications to the Senate Select Committee on Foreign Interference through Social Media that more time would be needed to assess the code's effectiveness.[26]

India

In the wake of a spate of vigilante attacks by mobs acting on rumours spread on WhatsApp, **the Indian Government has focused its legal response to harmful content on changes to the digital intermediary liability regime**, with the accompanying rhetoric centering on platforms falling short of putting in place adequate technological solutions and with complying with law enforcement agencies.[27]

On 26 July 2018, then-Minister of Electronics and Information Technology Shri Ravi Shankar Prasad responded to a Calling Attention motion on the misuse of social media platforms and the spreading of fake news, leading to rising incidents of violence and lynching in the country in the Rajya Sabha, the upper house of India's National Parliament.[28] He conveyed the Government's awareness of social media platforms being abused as vehicles "to commit crime, incite hatred, provoke terrorism, extremism, promote money-laundering, etc.", and indicated that while WhatsApp had responded to the Government's notice to put in place measures to prevent misuse of its platform, they were "not adequate to meet the challenges of the situation".[29] He emphasized that platforms "cannot evade its respon-

[22] Ibid., p. 370-372.

[23] ACMA, "Position paper", p. 10.

[24] Ibid., p. 4.

[25] Digital Industry Group Inc., Disinformation Code, February 22, 2021.

[26] Parliament of Australia, *Official Committee Hansard, Senate, Select Committee on Foreign Interference through Social Media, Friday, 30 July 2021*, July 30, 2021, p. 32-39.

[27] Annie Gowen, "India's Supreme Court warns of 'mobocracy,' urges government to pass anti-lynching law after deadly attacks", *The Washington Post*, July 17, 2018.

[28] Per the Parliament of India, Rajya Sabha, Council of States' *Practice & Procedure-Abstract Series*, a Calling Attention motion combines the asking of a question for answer with supplementaries and short comments in which different points of view are expressed concisely and precisely, and the Government has adequate opportunity to state its case. It gives members an opportunity to bring to the surface the failure or inadequate action of Government on a matter of urgent public importance.

[29] Parliament of India, Rajya Sabha, Council of States, *Parliamentary Debates Rajya Sabha Official Report Vol. 246, No. 7, Floor Version*, July 26, 2018, p. 457-459.

sibility, accountability and larger commitment to ensure (it) was not misused on a large scale to spread incorrect facts projected as news and designed to instigate people to commit crime", and unveiled the Government's proposal to reinforce the legal framework such that "if (the platforms) do not take adequate and prompt action, then the law of abetment also applies to them".[30] Prasad also briefly drew attention to the "weaponisation of information against India's strategic interest and economic stability", citing the possibility of the data of Indians being illegally obtained and misused in the case of Cambridge Analytica.[31] In response, Members of Parliament further indicated that political parties and activists often wielded false information to denigrate opponents and conducted "politics of hate".[32] This is an apt reflection of the Indian political climate, where disinformation has become politicized and political parties accuse one another of propagating "fake news" while denying wrongdoing on their part, despite documented evidence of associated "cyber troops" across political parties undertaking disinformation as part of political campaigning.[33]

The Government subsequently released in December 2018 the draft of what has become the *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 ("IT Rules, 2021")*. These are subsidiary rules which the central government exercised power delegated under Section 87 of the *Information Technology Act, 2000 (IT Act)* to enact, and which were notified under Section 79(2) of the same Act to specify the due diligence requirements to be observed by intermediaries to claim safe harbour protection for content hosted on their platforms. The "self-regulatory framework" is thus premised on the existing laws and statutes of the country, even as it seeks to impose new obligations to increase the accountability of intermediaries with respect to its terms of service and compliance to the rules.[34] Among other requirements, intermediaries with more than 5 million users are now required to proactively monitor and filter unlawful content, as well as provide for the traceability of users "in the interest of the sovereignty or integrity of India, defence of India, security of the State, friendly relations with foreign States or public order or for preventing incitement to the commission of any cognizable offence relating to above or for investigation of any offence".[35] The *IT Rules, 2021* also expands the powers to block information in case of emergencies to "persons" and "publishers", in addition to "intermediaries" which was already provided for under Section 69(A) of *the IT Act*.[36]

The *IT Rules, 2021* have since been invoked on two occasions – to block 20 YouTube channels and two websites allegedly part of a Pakistani coordinated disinformation operation; and to suspend over 75 social media accounts across Twitter, YouTube and Facebook for pushing "fake/inciting content" concerning the Government and targeting Hindu

[30] Ibid., p. 461.

[31] Ibid., p. 458-459.

[32] Ibid., p. 478.

[33] University of Oxford, Oxford Internet Institute, Programme on Democracy & Technology, *Country Case Studies Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation*, January 13, 2021, p. 177-182.

[34] PIB, "Government notifies".

[35] The Gazette of India, *Ministry of Electronics and Information Technology Notification, G.S.R. 139(E): the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021*, February 25, 2021.

[36] Ibid.; Section 69 of India Code, *The Information Technology Act, 2000*, June 9, 2000.

women.[37] It should be noted that there are currently no direct provisions in Indian law that specifically deal with "fake news", and "fake news" remains undefined under any domestic legal mandate, even as it has been defined by the Government.[38] Nonetheless, several offences in the *Indian Penal Code, 1860,* the *IT Act* and the *Disaster Management Act, 2005* criminalize certain forms of speech that may constitute "fake news", including the crimes of sedition, promoting enmity between different groups, and upsetting public tranquillity. Several laws, such as the *Code of Criminal Procedure, 1973* and the *Indian Telegraph Act, 1885* also support communication blockades, which India frequently resorts to in order to deal with fake news in the interest of maintaining public safety and averting public emergency.[39]

Industry-led efforts have failed to gain momentum. In February 2020, a grouping of digital platforms and publishers, fact checkers, civil society and academia, led by the Internet and Mobile Association of India, collectively named the Information Trust Alliance (ITA), was reported to have come together to control the spread of harmful content, including fake news.[40] While the grouping had reportedly discussed a Code of Practice to implement a centralised complaint registration mechanism and a standardised redressal process for disputed content, it was unable to reach a consensus given the different practices followed by the various platforms.[41]

New Zealand

Like Australia, New Zealand's existing legal framework has approached the issue of harmful and false information under the threat of foreign interference, albeit even more narrowly to only during electoral periods. Amendments to the *Electoral Act* – which already provided that a person is guilty of a corrupt practice if they deliberately publish a false statement with the intention of influencing an elector, on election day or the two days prior – were passed in December 2019 to increase the transparency of election advertisements, which must henceforth include name and address details across all mediums. Then-Justice Minister Andrew Little described the threat as "an avalanche of fake news social media ads that contain no information about who is behind them" with the aim of interfering with democracy, which had been observed in other countries, and which should not be repeated in New Zealand.[42] The amendments were passed under urgency, ahead of the conclusion of the Inquiry into the 2017 General Election and 2016 Local Elections conducted by the Justice

[37] PIB, India dismantles Pakistani coordinated disinformation operation, December 21, 2021; Aashish Aryan, "Over 75 accounts of various social media websites blocked for fake, inciting content", *The Indian Express*, January 9, 2022.

[38] In his detailed statement to the Rajya Sabha on 26 July 2018, Prasad defined "fake news" as a type of propaganda that consists of deliberate misinformation or hoaxes; spread in order to damage an agency, entity, or person, and/or gain financially or politically, create disturbance and unrest; often using sensational, dishonest, or outright fabricated headlines to increase readership, online sharing, and Internet click revenue.

[39] Matt Burgess, "To fight fake news on WhatsApp, India is turning off the Internet", *WIRED*, October 18, 2018.

[40] Megha Mandavia, "Social media to join hands to fight fake news, hate speech", *The Economic Times*, February 19, 2020.

[41] Ibid.

[42] New Zealand Government, *Government to ban foreign donations*, December 3, 2019.

Committee, which Little criticised as being released too late to allow additional safeguards to be implemented ahead of the 2020 General Election.[43]

Substantial work in this area is in progress, and there appears to be considerable appetite to introduce additional regulatory interventions. Amongst its recommendations, the Justice Committee proposed for election advertisements to be controlled even more tightly, notably by prohibiting foreigners from advertising in social media to influence election outcomes and ensuring enforceability through "appropriate constraints and legal obligations" on social media platforms.[44] More notably, it also singled out the issue of astroturfing, which was described as "the spreading of disinformation by robot accounts and paid participants, to give the appearance that a campaign has grass-root support", in response to intelligence agencies' indication that the New Zealand public was likely to encounter them given the international nature of online content, even though the country had not been the direct target of widespread, state-backed disinformation or malinformation campaigns in the 2017 election.[45] To address disinformation spread by foreign entities, the Justice Committee recommended the creation of an independent regulator with statutory powers to monitor technology companies and ensure they comply with a compulsory Code of Ethics, as well as greater clarity regarding the legal framework around social media services and their designation as "platform" or "publisher", drawing inspiration from the UK House of Commons' Digital, Culture, Media and Sport Committee's report on disinformation and 'fake news' and the Australian Joint Standing Committee on Electoral Matters' Report on the conduct of the 2016 federal election and matters related thereto.[46]

The Government has since indicated that the recommendations relating to electoral financing would be considered as part of a comprehensive review of electoral law following the 2020 General Election, while broader work on the regulation of social media platforms would be taken into account as part of a media content regulatory review led by the DIA.[47] Per the Classification Office, the threat of false and harmful information has also begun to gain public mindshare, ostensibly in view of the infodemic which accompanied the COVID-19 pandemic, as well as politics in the US.[48] Accordingly, it has recommended a broad strategy to address the issue which includes "updating regulation where needed" and putting in place industry agreements or Codes of Practice to create "a more consistent set of expectations and approaches" to moderation and industry responsibility.[49]

It is worth noting that New Zealand's legislative measures regarding harmful information online have thus far included, if not prioritised, tackling information which may not be

[43] Boris Jancic, "Probe into foreign interference too late: Justice Minister Andrew Little", *NZ Herald*, December 10, 2019.

[44] New Zealand Parliament, *Inquiry into the 2017 General Election and 2016 Local Elections: Report of the Justice Committee*, December 10, 2019, p.59.

[45] Ibid., p. 17.

[46] Ibid., p. 58-59.

[47] New Zealand Parliament, *Government Response to Justice Committee report: Inquiry into the 2017 General Election and 2016 Local Elections*, May 1, 2020.

[48] Classification Office, *The Edge of the Infodemic: Challenging Misinformation in Aotearoa*, June 2021, p. 15. The Classification Office is an independent entity responsible for censorship and classification of publications in New Zealand.

[49] Ibid., p. 52-53.

false. This includes information enabled by digital technology which causes "serious emotional distress" at the individual level, such as cyber bullying and digital harassment, as well as the livestreaming and distribution of "objectionable content" such as terrorist and violent extremist content, in the wake of the 15 March 2019 terrorist attacks in Christchurch, New Zealand, which saw the first instance of a killer live streaming his attack and which had gone viral.[50] To this end, the *Harmful Digital Communications Act* was passed in 2015 and *the Films, Videos, and Publications (Urgent Interim Classification and Prevention of Online Harm) Amendment Act* became law in February 2022. Both pieces of legislation include obligations on online hosts to take down or disable public access to prohibited content and establishes platforms' liability for failing to act on takedown notices, levers which could potentially eventually be adopted to tackle mis/disinformation.

Singapore

Singapore is most mature in its suite of legal responses in that it is the only one which has successfully enacted dedicated statutes to counter both deliberate online falsehoods, as well as information campaigns which are of foreign provenance and which are executed with hostile intent. Policymakers have placed the possibility of regulating the manipulation of information on the public agenda since April 2017, when Minister for Law and Home Affairs K Shanmugam underscored the global trend of the weaponisation of information by foreign actors to destabilise societies, and of private actors profiteering from fake news – which at times precipitate real world consequences, such as alarm to the public or the diversion of emergency resources, if not quickly corrected – citing occasions of the latter having already occurred in Singapore.[51] The discourse perpetuated by policymakers on the risks to Singapore has been largely premised on its attractiveness as a target given its openness and connectedness as a strategic node for international finance, trade, travel and communications, as well as the susceptibility of the city-state's racial and religious fault lines being easily exploited, the impact of which upends ethnic-religious harmony, undermines public institutions, interferes in elections and other democratic processes, and threatens the sovereignty of the country.[52] The successful positioning of the issue as an existential threat to the core tenets of the Singapore society and nation has been credited by some as the reason for the relatively unimpeded process in which the authorities were able to put in place legislative measures to contend with the threat.[53]

The ad hoc Select Committee on Deliberate Online Falsehoods – Causes, Consequences and Countermeasures, appointed in January 2018, reached similar conclusions after reviewing 170 written representations and hearing oral evidence from 65 individuals and

[50] Section 4 of the Harmful Digital Communications Act 2015; Section 132C of ilms, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill.

[51] Ministry of Law, *Oral Answer by Minister for Law, Mr K Shanmugam, to Parliamentary Questions on Fake News*, April 3, 2017.

[52] Ministry of Law, *Deliberate Online Falsehoods: Challenges and Implications, A Green Paper by the Ministry of Communications and Information and the Ministry of Law*, January 5, 2018, p. 16-19.

[53] Ric Neo, "The securitisation of fake news in Singapore", *International Politics*, 57, no. 4 (2019), p. 730.

organisations. It pushed for new legislation to be implemented to complement non-legislative measures, which alone were deemed insufficient to deal with the "strength and serious consequences" of deliberate online falsehoods.[54] The Select Committee recommended that the legislation provide "the necessary scope, speed and adaptability to deal with the realities of the phenomenon" to allow "calibrated" government intervention to counter the broad spectrum of falsehoods which are against the "public interest".[55] The result was the passing of the *Protection from Online Falsehoods and Manipulation Act* (*POFMA*) in October 2019, wich was designed as a graduated and proportionate tool with a spectrum of measures which would actually provide the Government narrower powers than it already has in existing laws, such as the issuing of "Correction Directions" to both individuals and intermediaries which requires a correction notice to be carried alongside a disputed online statement to allow readers to make informed judgments and draw their own conclusions.[56] Its broader purpose is to prevent communication of false statements; suppress online locations that repeatedly communicate false statements; enable measures to detect, control, and safeguard against coordinated inauthentic behaviour; and enhance disclosure of information concerning paid political content. It should be noted that a Correction Direction may be issued to a person even if he or she does not know or has reason to believe that the statement is false, and that penalties may apply only in the event of non-compliance with the direction.[57]

As of December 2021, *POFMA* has been used 33 times to date, more than half of which to deal with falsehoods concerning COVID-19, with the remaining falsehoods concerning suggestions that the Government is mismanaging public funds, abusing police power, favouring foreigners over locals, and carrying out judicial executions in an unlawful, brutal manner, among others.[58] All 33 occasions involved Correction Directions.[59] In addition, four Facebook pages were made "Declared Online Locations" (which entails the online location informing end-users that it is the subject of multiple active *POFMA* directions), and were subsequently the subject of "Access Blocking Orders".[60] One website was the subject of an Access Blocking Order for non-compliance with a Correction Direction.[61]

A second law has since been passed in October 2021, *the Foreign Interference (Countermeasures) Act (FICA)*, to specifically target foreign meddling in Singapore's domestic politics conducted through hostile information campaigns (HICs) online, described as "covert, coordinated and sophisticated online activities [which] seek to advance the interests

[54] Parliament of Singapore, *Report of the Select Committee on Deliberate Online Falsehoods – Causes, Consequences and Countermeasures*, September 19, 2018, p. 164.

[55] Ibid., p. 164-165.

[56] Per the *Second Reading Speech by Minister for Law, K Shanmugam on The Protection from Online Falsehoods and Manipulation Bill* on 7 May 2019, the Government already had the power to take down any objectionable material in the public interest under *the Broadcasting Act*, *the Telecommunications Act* and various other pieces of legislation.

[57] Section 11(4) and Section 15 of Protection from Online Falsehoods and Manipulation Act 2019.

[58] Kenny Chee, "Singapore's fake news law used 33 times to date, including 19 against Covid-19 misinformation", *The Straits Times*, December 1, 2021; Aqil Haziq Mahmud, "IN FOCUS: Has POFMA been effective? A look at the fake news law, 1 year since it kicked in", *Channel NewsAsia*, October 3, 2020.

[59] Ministry of Communications and Information, *MCI's response to PQ on Breakdown of Number of Declarations, Directives, Orders and Notices Issued under Protection from Online Falsehoods and Manipulation Act to-date by Ministries*, November 3, 2021.

[60] Ibid.

[61] Ibid.

of the attacking country".[62] Besides making it an offence for individuals to conduct clandestine foreign interference by electronic communications activity, *FICA* introduces powers for targeted and calibrated directions to allow the Government to achieve three objectives – namely, to obtain information about HICs, to detect and prevent HICs from taking place, and should they occur, to contain the HIC.[63] To this end, the Minister for Home Affairs can exercise extraterritorial jurisdiction to oblige global entities such as social media platforms to assist in investigating and preventing such potential HIC threats, as "(Singapore) cannot wait for harms to occur before taking action, because severe damage may already been done".[64] *FICA* also provides for other directions reminiscent of *POFMA*, such as those requiring digital service providers to expose HICs by carrying mandatory messages from the Government to warn Singaporeans of such HICs, as well as disabling public access to content or a particular social media or electronic service that is part of such campaigns.

Taiwan

In the case of Taiwan, its key preoccupation can be attributed to China's reported growing insistence for "complete reunification", against a backdrop of disinformation efforts allegedly mounted by Chinese actors to shape the domestic political narrative, shoring up negative sentiments towards independence-leaning President Tsai Ing-wen, and creating an image of an incompetent leadership.[65] Tsai has not been shy about declaring the pressing need to actively deal with disinformation to defend Taiwan's free and open democratic society from a specific hostile external force. In her 2018 National Day address, Tsai espoused the need to prevent "foreign powers from infiltrating and subverting (Taiwan's) society, ensuring that (Taiwan's) democratic institutions and social economy function normally", and identified the "systematic dissemination of disinformation" as a central means of doing so, placing the threat alongside more traditional security threats such as military coercion, diplomatic pressure and predatory economic policies.[66] Shortly before, Taiwanese media reported that the Ministry of Justice's Investigation Bureau, which established a big-data and public opinion task force, found "unequivocal evidence" that Beijing was responsible for spreading fake news articles with the objective of manipulating public opinion in Taiwan.[67] It was also reported that the National Security Bureau had briefed the Legislative Yuan's Foreign Affairs and National Defense Committee that China was "behind a propaganda campaign to interfere with Taiwan's elections by creating disinformation and fake

[62] MHA, *First Reading of Foreign Interference (Countermeasures) Bill*, September 13, 2021.

[63] MHA, *Second Reading of Foreign Interference (Countermeasures) Bill – Speech by Assoc Prof Muhammad Faishal Ibrahim, Minister of State, Ministry of Home Affairs and Ministry of National Development*, October 4, 2021.

[64] Ibid.

[65] Huaxia, "Full Text: Speech by Xi Jinping at a ceremony marking the centenary of the CPC", *Xinhua News Agency*, July 1, 2021; Scott W. Harold, Nathan Beauchamp-Mustafaga, and Jeffrey W. Hornung, *Chinese Disinformation Efforts on Social Media*, RAND Corporation, 2021, p. 66; Sean P. Quirk, "Lawfare in the Disinformation Age: Chinese Interference in Taiwan's 2020 Elections", *Harvard International Law Journal* 62, no. 2 (Summer 2021), p. 530-540.

[66] Office of the President, Republic of China (Taiwan), *President Tsai delivers 2018 National Day Address*, October 10, 2018.

[67] Li-chung Chien, Li-hua Chung and Jonathan Chi, "China using fake news to divide Taiwan", *Taipei Times*, September 16, 2018.

news targeting Taiwanese media outlets, radio and television programs and Web sites".[68] It is thus unsurprising that the Legislative Yuan passed *the Anti-Infiltration Act* about two weeks before the 2020 presidential and legislative elections. The Act's stated objectives are to "strengthen mechanisms to defend Taiwan's democracy, simplify cross-Strait exchanges, and restore order in interactions between the two sides", and it specifically criminalizes the spreading of disinformation to interfere with elections on behalf of hostile external forces.[69]

That said, the authorities have not ignored the wider sources and harms of disinformation. From late 2018, it undertook a sweeping review of its regulatory system and has since introduced amendments to at least eight existing pieces of legislation. It makes it an offence to disseminate rumours or false information relating to a range of issues, from outbreaks of communicable diseases (under *the Communicable Disease Control Act*), nuclear accidents (under *the Nuclear Emergency Response Act*) to anything that could affect market food prices (under *the Food Administration Act*), among others.[70] The review has notably also included amendments which impose harsher penalties for disinformation spread via mass communication tools, in acknowledgement of the more extensive and serious harm caused by the rise of new internet technologies and social media. *The Social Maintenance Order Act* has commonly been invoked to punish the dissemination of disinformation. The number of cases has increased over the years, from 21 cases in 2018 to 151 cases in 2019 and 233 cases between January and May 2020.[71] However, conviction rates are low (with a non-punishment rate of 72.2% in 2019), and even when they occur, rarely lead to significant penalties like prison terms or steep fines.[72]

Beyond the aforementioned criminal sanctions against purveyors of disinformation, the National Communications Commission (NCC) has also drafted a *Digital Communications Bill*, which is intended to better regulate "inappropriate" content on online platforms by holding operators accountable for the content shown on their platforms, establishing a one-stop service for people to file complaints over online content, obliging large international platforms to participate in a self-regulatory mechanism to address international issues such as Internet violence and fraud, and authorising the NCC to work with "overseas partners" to address problematic content originating overseas.[73] Tsai's Democratic Progressive Party (DPP), despite having a solid majority in the Legislative Yuan, has not been able to advance the Bill since the first draft was submitted to the Executive Yuan in April 2017. Opposition political parties and global internet and technology companies alike have raised concerns regarding the NCC's executive overreach and the impact of the act on freedom of speech.[74]

[68] Li-hua Chung and William Hetherington, "China targets polls with fake accounts", *Taipei Times*, November 5, 2018.

[69] Mainland Affairs Council, Republic of China (Taiwan), *Legislative Yuan Passes Anti-Infiltration Bill to Strengthen Defense for Democracy and Preserve Stable and Orderly Cross-Strait Exchanges*, December 31, 2019.

[70] Taiwan High Prosecutors Office, *Anti-disinformation Policy Overview 2019*, May 16, 2019, p. 23-28.

[71] The Control Yuan – Republic of China (Taiwan), 監察委員新聞稿 [Press Release of the Supervisory Committee], 9 July 2019; Freedom House, *Freedom on the Net 2021, Taiwan*, accessed February 25, 2022.

[72] Ibid.

[73] Sherry Hsiao, "NCC proposing new act to censor Internet: KMT", *Taipei Times*, December 15, 2020; Shelly Shan, "NCC outlines digital act progress", *Taipei Times*, May 7, 2021.

[74] Hsiao, "NCC proposing new act"; Yun Chen and Jake Chung, "TPP questions NCC's draft digital act", *Taipei Times*, December 30, 2020; Asia Internet Coalition, *AIC Submits Industry Response on Taiwan's Draft Digital Communications Act (12 December 2018)*, December 12, 2018.

In response, the authorities have said they have "no choice but to put controlling regulations in place" given the "serious nature" of the problem, while underscoring that "any measures aimed at preventing harms from misinformation can only be undertaken on the condition that (freedom of speech) is upheld".[75]

On the industry front, the Taipei Computer Association and five major internet platforms in Taiwan – Facebook, Google, LINE, Yahoo, and Professional Technology Temple (PTT) – also appear to have ramped up their joint efforts in countering disinformation, and drew up self-regulatory principles regarding false information disseminated via their platforms in June 2019.[76] The *Industry Code of Practices for Misinformation Self-Regulation* (不實訊息防制業者自律實踐準則) is based on the Manila Principles of Intermediary Responsibility, and comprises four key commitments and 13 practical guidelines to help the public identify false information, including cooperation with third parties and the authorities to establish and maintain an independent, transparent and impartial monitoring mechanism.[77] There has been limited information on the effort since the initial announcement, with only Facebook releasing a report on its efforts to defend election integrity during the 2020 presidential and legislative elections.[78]

A COMPARATIVE ANALYSIS OF THE LEGISLATIVE INSTRUMENTS

While academics have recognised that legislation should be considered as part of a multipronged strategy to combat the threat of false and harmful information, there is a lack of studies on its effectiveness as a countermeasure.[79] This is likely because initiatives to enact and implement such legislation are generally at a nascent stage, and it is premature to offer a holistic impact assessment of both its intended and unintended effects. Any effectiveness evaluation would require deep analysis and data to consider its necessity, the proportionality, the transparency, accountability and due process from the competent authority charged with implementing the law, as well as the impact it has had on the behaviour of digital platforms and the people using them – all of which are context-sensitive. Even as some of these effects become observable over time, others such as the crime prevention effect of the threat of punishment, and the mitigating impact on the influence of false and harmful information remain challenging to measure. Nonetheless, the composite picture that arises from the previous section offers several insights into regulatory trends taking shape.

First, the effectiveness in reducing or eliminating the adverse impacts of fake news does not appear to be the principal determinant of the primary regulatory response invoked, nor does normative compliance with the applicable rules of international

[75] Executive Yuan, *Measures to counter misinformation predicated on preserving free speech*, December 12, 2018.

[76] Taipei Computer Association (TCA), 自律先行_本會與四大平台業者攜手防制不實訊息(2020年自律成果報告陸續更新上架), June 21, 2019. PTT is Taiwan's version of Reddit.

[77] Ibid.

[78] TCA, *Defending Election Integrity in Taiwan*, October 6, 2020.

[79] Gulizar Haciyakupoglu et al., "Countering Fake News: A Recent Survey of Global Initiatives", S. Rajaratnam School of International Studies, March 2018, p. 20-21.

human rights laws. When weighed against these two parameters, existing literature contends that criminal sanctions are the most effective regulatory response. Its deterrent effect targets the initial creation and sharing of the false news, such that the public are never exposed to associated misinformation or disinformation, while the criminalization of false news in a context-specific fashion can be justified and should not be dismissed on account of broad normative claims to protect the freedom of expression.[80] Comparatively, evidence-based psychological analysis demonstrates that information correction measures and content removal or blocking measures are less effective at mitigating the effects of false news once it has been created, and can even be harmful due to potential 'backfire' effects.[81] From a normative perspective, information correction measures are the least intrusive, while content removal or blocking measures are the most intrusive.[82] **Why Taiwan, India and Singapore have in recent times defaulted to different legislative responses**, viz. criminal sanctions, blocking measures, and information correction measures respectively, **could be due to their respective unique social context and circumstances, for which the stated legislative intents offer some insight.**

Indeed, Taiwan's benchmark of effectiveness appears most aligned to the abovementioned considerations alone, with the Taiwanese authorities being unsurprised and in fact encouraged by the low rate of convictions, interpreting it as a reflection of the judiciary's narrow interpretation of what it means to spread rumours "in a way that is sufficient to undermine public order and peace", and in doing so, is playing its role in ensuring that freedom of expression is not unjustifiably infringed upon.[83] On the other hand, Singapore's decision to rely primarily on information correction could be explained by its stated intent to combat "low level everyday falsehoods" which do not always cause an immediate impact but could foment long-term societal damage by skewing world views over time.[84] Facilitating content discovery and access to different sources representing diverse perspectives thus appears more suited to this objective. Finally, India's tendency to resort to blocking measures is likely informed by the historical propensity of false information to become viral and incite mob violence in the country, and thus the need for a more intrusive intervention of disabling access to and arrest the propagation of false information to maintain public order. Without entering into a debate on whether the policy agenda of pursuing legislation to counter false and/or harmful information is justified, or the suitability of any particular tool against the relevant legal threshold for their specific contexts, this observation provides an understanding of the utility of the laws in place and shows that the assessed suitability of a particular regulatory response is highly context-specific.

Second, laws targeted at foreign interference have been rarely invoked. A widely-reported case in 2020 involving a social media group alleged to be acting on behalf of

[80] Rebecca K Helm and Hitoshi Nasu, "Regulatory Responses to 'Fake News' and Freedom of Expression: Normative and Empirical Evaluation", *Human Rights Law Review* 21, No. 2 (June 2021), p. 323-325.

[81] Ibid., p. 315-322.

[82] Ibid., p. 315-322.

[83] The Control Yuan – Republic of China (Taiwan), 監察委員新聞稿 [Press Release of the Supervisory Committee], 9 July 2019; Shih-Shiuan Kao, "Taiwan's Response to Disinformation: A Model for Coordination to Counter a Complicated Threat", The National Bureau of Asian Research, September 16, 2021.

[84] Parliament of Singapore, "Select Committee", p. 162.

the Chinese state apparatus to encourage New South Wales Labor Member of Parliament Shaoquett Moselmane to champion Chinese Government interests has demonstrated how foreign influence is challenging to prosecute. Even though the Australian Federal Police has searched the homes and offices of Moselmane and his policy advisor John Zhang, the Australian Border Police has intercepted Zhang's communications with China's diplomats, and the ASIO has questioned Chinese-Australian academics and journalists who were part of the social media group, no penalties have been sought yet.[85] This could be due to the difficulties in establishing guilt "beyond reasonable doubt" – the highest legal threshold in common law systems – and the challenge faced by the authorities in mounting a criminal case against those suspected of acting on behalf of foreign principals, even after the diplomatically problematic decision to accuse another country publicly had been taken.[86]

Another example which points to how difficult it is to determine the actual scope and intent – much less efficacy – of influence campaigns is the September 2018 incident, where Taiwanese Diplomat Su Chii-cherng who was stationed in Osaka, Japan, committed suicide after being falsely accused on social media of failing to evacuate Taiwanese tourists stranded in the wake of Typhoon Jebi. While many observers have drawn the conclusion that the online rumour originated from China, there is less agreement concerning its objective, with some postulating that it was intended as a Chinese-made but Taiwanese-fuelled disinformation campaign in which the rumour was planted by China in the hopes that it would be picked up by unwitting Taiwanese media focused more on getting more eyeballs than on protecting information integrity, while others argued that it was intended as a domestic propaganda campaign in view of the story having gone viral on Weibo amongst Chinese mainlanders first.[87]

Almost three years after the incident, two Taiwanese who had insinuated Su's culpability through a PTT post and hired internet trolls to criticize Su's office were sentenced in November 2021 to 6 months of imprisonment for insulting the public office under *the Criminal Code of the Republic of China*.[88] Alas, the incident pre-dates Taiwan's anti-disinformation laws including *the Anti-Infiltration Ac*t, and the courts have not had to resolve the question of intent. While the conviction serves to profile how spreading false information does precipitate real world harms and could act as a deterrence for would-be purveyors of online rumours, it can be argued that this is ultimately a red herring in an incident that Taiwanese leaders, including President Tsai, have held up as a symbolic example of the threat posed to Taiwan by Chinese disinformation operations.[89]

[85] Rubinsztein-Dunlop and Hui, "Australian police"; Tony Walker, "What can Singapore learn from Australia's foreign interference countermeasures?", *South China Morning Post*, October 4, 2021.

[86] Katherine Mansted, "The Domestic Security Grey Zone: Navigating the Space between Foreign Influence and Foreign Interference", National Security College, The Australian National University, February 2021, p. 7-8.

[87] Stephen J. Hartnett and Chiaoning Su, "Hacking, Debating, and Renewing Democracy in Taiwan in the Age of "Post-Truth" Communication", *Taiwan Journal of Democracy* 17, no. 1 (July 2021), p. 39; Bo Julie Crowley, Casey Cocoran and Raina Davis, "Disinformation Threat Watch: The Disinformation Landscape in East Asia and Implications for US Policy", Belfer Center for Science and International Affairs, Harvard Kennedy School, May 2019, p.19.

[88] Jane Lee, "Yang Huiru Sentenced to 6 Months on Charges of Insulting Public Office", *International Community Radio Taipei (ICRT)*, November 12, 2021.

[89] Nick Aspinwall, "Taiwan's War on Fake News is Hitting the Wrong Targets", *Foreign Policy*, January 10, 2020.

Third, even as strikingly different approaches have been adopted to counter false and/or harmful information, **certain aspects and underlying considerations of governance have prevailed, most notably in the increased allocation of responsibility to digital platforms**. This is demonstrated by the enactment of the Australian *Code*, India's *IT Rules, 2021*, and Singapore's *POFMA* and *FICA*. That said, the "how" is still up for debate, as reflected by the fact that these are distinct regulatory instruments which allow for different degrees of enforcement – or encouragement – of compliance by digital platforms. The Australian *Code* is a **co-regulatory framework**, where an independent regulator appointed by the government has provided guidelines to steer the industry's development of the code, but which **cannot be enforced in a court of law**; India's *IT Rules, 2021* is **subordinate legislation**, drafted by the executive, and is **only enforceable when it is within the ambit of what is permitted under its parent law** (which provides the power to issue directions to block public access to as well as intercept, monitor or decrypt information if it is necessary or expedient to do so in the public interest); while Singapore's *POFMA* and *FICA* are **traditional regulation** drafted by state bureaucrats. *POFMA* also provides for subordinate legislation in the form of Codes of Practice issued by the competent authority.[90]

The ACCC's preference for co-regulation through an industry-drafted code was to better ensure that stakeholder concerns and practical considerations, such as cost of compliance were managed, and to avoid the direct involvement of the Government to balance public interest with free speech and the right of individuals to choose.[91] While not explicitly communicated by the Indian Government in relation to the *IT Rules, 2021*, per the Rajya Sabha, the case for subordinate legislation in general pertains to its flexibility, elasticity, expedition and opportunity for experimentation, given that it allows for the executive, experts and technocrats to provide for working details within the framework of legislation which need not be debated in Parliament.[92] In this manner, subordinate legislation supports the future-proofing of regulation as it can be more adaptable and flexible to the rapid changes that characterize the digital age, by overcoming the often lengthy legislative process of enacting new laws. That said, the Indian Government has already run into implementation difficulties, with several petitions having been filed in the high courts of Bombay, Calcutta, Delhi, Madras, Karnataka and Kerala challenging various aspects of the *IT Rules, 2021,* including its overreach and constitutional validity.[93] Singapore's approach, while leveraging conventional means of the state apparatus, has not been spared of criticisms. The right to freedom of expression being undermined is often cited by Western NGOs and media, and local politicians and journalists have also expressed similar concerns.[94] The authorities have nonetheless continued to defend the

[90] Three Codes of Practice have been issued to date – *Code of Practice for Giving Prominence to Credible Online Sources of Information*, *Code of Practice for Transparency of Online Political Advertisements*, and *Code of Practice for Preventing and Countering Abuse of Online Accounts*.

[91] ACCC, "Platform Inquiry", p. 372.

[92] Parliament of India, Rajya Sabha, Council of States, Committee on Subordinate Legislation, February 2005.

[93] The Indian Express, *SC tags Centre's plea for transfer of petitions on new IT Rules from HCs with pending matter*, July 9, 2021; Dodhiya Kay, "Petition in Bombay HC challenges Information Technology Rules, 2021", *Hindustani Times*, July 4, 2021; Internet Freedom Foundation (IFF), *Table summarizing challenges to IT Rules, 2021 pending before High Courts*, December 9, 2021.

[94] Jeangène Vilmer, "Information Defense", p. 9-10.

law, with the Singapore Court of Appeal passing a landmark decision in October 2021, which upheld its constitutionality and rejected assertions that there is no judicial oversight over the exercise of powers under the law.[95]

The merits of the different regulatory instruments remain to be evaluated as they mature and enforcement of them expands – a fact not lost on the Australian authorities who have from the outset considered the possibility of more "direct" and "significant" regulation eventually replacing or supplementing the *Code* depending on the latter's effectiveness, which can be interpreted as tacit acknowledgement that the lack of legislative teeth could pose a challenge to compliance or the drafting of a meaningful *Code*.[96]

COMMON FEATURES OF REGULATORY MEASURES AIMED AT PLATFORM GOVERNANCE

Enforcement mechanism aside, this section examines the Australian *Code*, India's *IT Rules, 2021*, and Singapore's *POFMA* and *FICA* in greater depth, to draw out the common features of the regulatory measures aimed specifically at platform governance to better counter false and/or harmful information. The view that effective governance of the harms caused by digital media platforms requires intergovernmental collaboration in formulating shared strategies and rules has been gaining traction, evinced by the growing number of countries sending their parliamentarians to participate in the International Grand Committee of Disinformation and "Fake News", an ongoing forum for information-sharing, collaboration, and harmonisation of policies amongst democratic states actively considering legislation to address digital threats.[97] The proposed taxonomy of the measures established by the three aforementioned jurisdictions could serve as a precursory contribution towards realising this view. By enumerating policies already applied to digital platforms, it serves both to indicate where convergence already exists, as well as highlights measures which could be transposed elsewhere. The New Zealand and Taiwan models have been excluded in this comparison, given that the former is still in the process of shaping their approach towards containing the dissemination of false and harmful information, while the latter has enacted new laws which provide solely for ex-post penalties to be imposed on perpetrators of disinformation at the individual level. As such, both offer limited room for comparison.

Taxonomy of Measures

The three jurisdictions appear to largely converge, with the specific measures in place resembling one another. Notably, they have all adopted a **differentiated approach**, with the *IT Rules, 2021* and the Codes of Practice issued under *POFMA* imposing additional

[95] Ian Cheng, "Court, not minister, makes final decision under POFMA on whether statement is true or false: AGC", *Channel NewsAsia*, October 9, 2021.
[96] ACCC, "Platform Inquiry", p. 34.
[97] International Grand Committee on Disinformation (IGC), *Who We Are*, accessed February 25, 2022.

obligations taking into account the size of intermediary, while the Australian *Code* specifies minimum commitments and opt-in measures to accommodate the different services and products provided by digital platforms. Using the Institute for Strategic Dialogue's categorisation of digital policy approaches, the regulatory instruments across the three jurisdictions also demonstrate elements of both a **systemic approach**, whereby online platforms must demonstrate that their policies, processes and systems are designed and implemented with respect to the potential negative outcomes that could occur across a range of possible harms, as well as a **content-based approach**, which focuses on the effective and timely removal of content, often targeting a specific online harm, such as hate speech or electoral disinformation.[98] More concretely, the measures are structured around several general themes, including:

Systematic Measures

- **Duty of care:** measures to ensure the safety and integrity of their services. These include (i) *publishing clearly defined rules and regulations, privacy policies, or user agreements*, such that users are made aware of the behaviours which are prohibited; (ii) *ensuring the responsible curation of information* presented to users, which could include ranking algorithms which demote false and/or harmful news and conversely promote content from authoritative sources or sources which meet certain editorial standards; and (iii) *measures to prevent and manage the abuse of online accounts*, such as user identification verification features to prevent inauthentic accounts, disclosure of the use of bots, and user reporting mechanisms.

- **Santa Clara Principles 1.0:** transparency and accountability measures around the moderation of user-generated content.[99] This has also been termed by some as **user-centric moderation practices**, in recognition of the risks of content moderation, particularly when relying on exclusively technological solutions which do not involve the man-in-the-loop.[100] This includes (i) *regular, publicly-accessible compliance reports*, which should contain information about the detection of and measures taken to manage content and behaviours in violation of platform policies; (ii) *providing notices to users who are the subject of moderation measures*, to explain the action being taken and the grounds for such action; and (iii) *providing redress mechanisms for users to dispute content moderation actions*, which in effect offer users the opportunity to present any additional information as well as for human review.

- **Duty of cooperation:** processes and structural elements to facilitate cooperation with the authorities to fight against the dissemination of false and/or harmful information. This includes (i) *establishing designated facilities, channels or points of contact*, to allow for ease and certainty of communication, particularly in the case of non-compliance by the digital intermediaries themselves; and (ii) *providing assistance for investigations*, such as providing information on the provenance of an account, technical information about

[98] Institute for Strategic Dialogue, Digital Policy Lab '20 – Companion Papers, April 13, 2021, p. 19.

[99] The Santa Clara Principles on Transparency and Accountability in Content Moderation, Santa Clara Principles 1.0, accessed February 25, 2022.

[100] Alaphilippe et al., "Automated tackling", p. 74-75.

an account's activity, or enabling the identification of the first originator of a piece of content.

- **Preserving the freedom of expression and information:** measures which aim to respect and protect the right to seek, receive and impart information and ideas. This includes (i) *providing users control over the information they receive*, such as the ability to access alternative sources of information or to opt out of exposure to certain types of digital content; and (ii) *ensuring the right of reply*, such as providing for corrections or alternative discourse to be placed alongside a contested piece of information.

Content-based Measures

- **Encouraging media literacy:** transparency measures to allow users to critically evaluate content they come across. This includes (i) *disclosing advertisements or sponsored content*, to enable users to understand that they have been targeted; (ii) *identifying the source of the content,* such as exposing the meta data of a content source or the source of political advertising*;* and (iii) *indicating the reliability of a source*, such as providing trust indicators of content or labelling online locations which have repeatedly produced or hosted content which are false and/or harmful.

- **Stemming the spread of false and/or harmful information:** measures to arrest the propagation of and user exposure to false and/or harmful behaviour already occurring. This includes (i) *ensuring that the content is no longer available to users*, such as by removing the content or disabling access to the content; (ii) *reducing the propagation by inauthentic behaviour*; such as by restricting or suspending the functionality of user accounts engaged in inauthentic behaviour; (iii) *alerting users who have encountered or might encounter false and/harmful information*, such as labelling false content and issuing notices to users who have encountered and/or may encounter a false piece of content; and (iv) *establishing techniques to proactively identify prohibited content and behaviour*, to reduce the propagation of and thus user exposure to false and/or harmful behaviour, such as technological tools to identify prohibited content, identical content previously taken down or inauthentic behaviours.

Others

- **Disincentivizing bad behaviour:** policies and processes that aim to disrupt monetisation incentives and sources of financial support for false and/or harmful information. This includes (i) *making involvement a punishable offence at the individual level*, such as receiving financial or material benefit for providing a service for the communication of false statements, and higher penalties for the use of bots; and (ii) *disgorging online locations that propagate false and/or harmful information*; such as by prohibiting persons from providing financial support to and restricting the availability of advertising services and paid placements on such online locations.

While the range of measures and intended outcomes are laudable, drawing upon the EU's experience with the implementation of the *Code of Practice on Disinformation*, it should be noted that a fundamental issue lies in the inability of the authorities to conduct an independent evaluation of the timeliness, comprehensiveness and impact of the digital platforms' actions.[101] The platforms themselves are currently tasked to "grade their own homework" through self-assessment compliance reports, and the authorities are thus beholden to their willingness to share information and data. The crux of the issue is captured by former Facebook employee turned whistle-blower Frances Haugen in her opening statement to the United States Senate Committee on Commerce, Science and Transportation in October 2021: "This inability to see into Facebook's actual systems and confirm they work as communicated is like the Department of Transportation regulating cars by only watching them drive down the highway. Today, no regulator has a menu of solutions for how to fix Facebook because Facebook didn't want them to know enough about what's causing the problems — otherwise there wouldn't have needed to be a whistleblower. How is the public supposed to assess if Facebook is resolving conflicts of interest in a way that is aligned with the public good if the public has no visibility into how Facebook operates?"[102] Indeed, digital advocacy groups have already noted the shortcomings of the monthly compliance reports published as required under India's *IT Rules, 2021*, including the use of misleading metrics and the lack of uniformity in reporting across platforms, which ultimately betray the intent of engendering greater transparency and accountability of platforms' policies on harmful information.[103] Ensuring more meaningful information disclosure by the digital platforms is thus a key aspect in determining the success of many of the above-mentioned regulatory measures, although how that can be achieved is outside the scope of this paper.

CONCLUSION

The trend towards legislating false and harmful information in the Asia-Pacific is unmistakable, with many new pieces of regulation targeting the creation, distribution and manipulation of false and harmful information being enacted in the last four years, and many more still being drafted and considered. The primary finding of this paper illustrates that the regulatory strategies in the region are fragmented, and are adapted to their perceived threats, as well as respective social context and circumstances, among other factors. Even as the increased allocation of responsibility to the digital platforms presents a common area of focus in the absence of successful industry-led efforts, the enforcement mechanism has also been approached in divergent ways. While the goal of achieving shared, effective strategies to address common challenges in the digital space appears elusive, a diversity of regulatory strategies is arguably an imperative step towards the harmonisation of policies, serving to offer instruction on what proves effective or otherwise. Ultimately, contextual circumstances coupled with the rapid changes that characterize the digital age necessarily mean that the fight against false and harmful information is ever-dynamic, and will not benefit from a one-time, one-size-fits-all solution.

[101] European Commission, *Assessment of the Code of Practice on Disinformation – Achievements and areas for further improvement*, September 10, 2020, p. 19.

[102] Frances Haugen, *Opening Statement to Senate Committee on Commerce, Science & Transportation*, October 5, 2021.

[103] IFF, *#SocialMediaComplianceWatch: An analysis of Compliance Reports for the month of October*, January 21, 2022.

Celine Tham is a visiting research fellow at IRSEM. She has held various responsibilities within the Singapore Ministry of Defense (MINDEF). As a media relations officer, she was responsible for external communication activities aimed at strengthening public trust in MINDEF and the Singapore Armed Forces (SAF), as well as enhancing its international standing. To this end, she supported MINDEF/SAF personnel in their role as spokespersons, and worked closely with the media to achieve desired public communications outcomes. More recently, she has worked on the medium to long-term build-up of the ministry's capabilities for effective strategic communications. She graduated from Nanyang Technological University in Singapore with a Bachelor of Communication Studies. The views expressed in this report are those of the author alone. They do not reflect the views of any institutions the author is affiliated with.

Contact: celine.tham@irsem.fr