Regulating online hate speech: The Singapore experiment

Chen Siyuan, Yong Pung How School of Law, Singapore Management University, Singapore, Singapore

Published in International Review of Law, Computers and Technology, 2023

DOI: 10.1080/13600869.2023.2295091

Abstract: In Singapore, the introduction of the Protection from Online Falsehoods and Manipulation Act 2019 (POFMA) generated considerable debate and feedback. One of the main concerns was whether the law would unduly restrict the freedom of expression. In focusing on the provisions of POFMA that could possibly apply to hate speech, this paper situates the criticisms within the larger framework of international human rights law and international practices and proposes some ways forward to improve the regulatory framework for online hate speech.

Keywords: Online hate speech, fake news, Singapore

1. Context

Internet-based social media platforms may have been around for a while, but governments continue to struggle to regulate online speech in a way that strikes an appropriate balance between freedom and security (Kaye 2018) – and which also accords with jurisdictional propriety (Chia Citation2018). The latest conundrum presented by social media in particular is hate speech. Hate speech is claimed to have the potential of disrupting society by sowing discord and division between different racial or religious groups, influencing political outcomes, legitimising fringe views, and discrediting establishment journalism (Posetti et al. 2018). Given the nature of the Internet, the spreading of such information is not confined to persons within a single jurisdiction, and can include persons from other jurisdictions and even bots, through coordinated means or otherwise.

Around the world, responses by governments to this phenomenon have ranged from proposed fines (Fox 2019) and imprisonment (Saudi Gazette Citation2018) to actual shutdowns of Internet services (Lever 2019). Singapore has responded principally through the introduction of the Protection from Online Falsehoods and Manipulation Act 2019 (now popularly known as POFMA), a piece of legislation that comprises no less than 60 sections (Tham 2019a).[1] During the Second Reading, Minister for Home Affairs Mr

K Shanmugam defined 'hate speech' as 'all forms of expression which spread, incite, promote, or justify racial hatred, xenophobia, or other forms of hatred based on intolerance' (Singapore Parliament 2019a). This was adopted from the Council of Europe Committee of Ministers' definition in 1997 (Council of Europe Committee of Ministers 1997; Ministry of Home Affairs 2019). Thus, while POFMA seems to cover more than just 'hate speech', the focus of this article will be on the effectiveness of POFMA as a tool in combatting hate speech.

POFMA has generated considerable debate and feedback since it was first mooted in 2017, be it in the form of the protracted Select Committee hearings on Deliberate Online Falsehoods,[2] the Green Paper on Deliberate Online Falsehoods, or the public consultations that have ensued after the First Reading of the bill (Ministry of Law 2019a). International groups such as the International Commission of Jurists (Tan 2021), Reporters Sans Frontiers (Reporters Without Borders 2020), and Human Rights Watch (Human Rights Watch 2021) have also weighed in, claiming that the law would unduly suppress speech and possibly be abused by the government against political opponents. David Kaye, a UN Special Rapporteur, even argued that POFMA 'raises serious concerns regarding freedom of expression', and that it would be 'incompatible with international human rights law' (Kaye 2019).

The objective of this paper threefold: first, to identify and analyse some of the issues presented by the key provisions of POFMA, with a particular focus on provisions relating to online hate speech; secondly, to situate these provisions within the framework of both international law and international practices; thirdly, to make recommendations on how else the law regulating online hate speech can be improved, as well as to make brief observations on the application of POFMA in the three years since its introduction. To be clear, POFMA's ambit extends beyond online hate speech, but those areas will not form the focus in this comment due to space constraints.

2. Examining POFMA provisions and identifying some of the possible criticism

We begin our analysis by examining the salient POFMA provisions that pertain to online hate speech, before setting out some of the potential issues:

a. Under section 7(1), a person must not do any act in or outside Singapore to communicate[3] a statement knowing or having reason to believe that it is a false statement of fact, and the communication of that statement is likely to, *inter alia*, prejudice national security,[4] public health, public safety, or public tranquillity, or incite feelings of enmity, hatred, or ill-will between different groups of persons.[5] As will be seen, the latter half of this provision (especially the part on incitement of feelings) essentially refers to the consequence of hate speech, while the former half refers to the act that causes it (a false statement of fact).[6]
b. Under section 7(2), a person guilty of section 7(1) shall be fined not exceeding $50,000 or be imprisoned for a term not exceeding 5 years, or both. If the person had used an inauthentic online account or a bot to commit the offence, he shall be fined not exceeding $100,000 or be imprisoned for a term not exceeding 10 years, or both.

c. Under sections 10, 11, and 12, a minister may order a correction direction or a stop communication direction if a false statement of fact has been communicated in Singapore and the minister is of the opinion that it is in the public interest to issue the direction. '[I]n the public interest' is defined non- exhaustively in section 4, and the definition that is of relevance here is the doing of something that is necessary or expedient: in the interest of the security of Singapore; to protect public health, public safety, or public tranquillity; or to prevent incitement of feelings of enmity, hatred, or ill-will between different groups of persons.
d. Under section 15, a person who, without reasonable excuse, fails to comply with an order made under sections 11 or 12 without reasonable excuse shall be fined not exceeding $20,000 or be imprisoned for a term not exceeding 12 months, or both.
e. Under sections 17 and 19, a person subject to an order made under sections 11 or 12 may appeal to the High Court after he has applied to the minister to vary or cancel the order, and the minister has refused the application in whole or in part.[7]
f. Under sections 20, 21, 22, and 23, a minister may order an internet intermediary to make a general correction, make a targeted correction, or disable access of end-users if a false statement of fact has been communicated in Singapore and the minister is of the opinion that it is in the public interest to issue the direction.[8]
g. Under section 27, a person who fails to comply with an order made under sections 21, 22, or 23 without reasonable excuse shall be fined not exceeding $20,000 or imprisoned for a term not exceeding 12 months, or both.
h. Under section 29, a person subject to an order under sections 21, 22, or 23 may appeal to the High Court after he has applied to the minister to vary or cancel the order, and the minister has refused the application in whole or in part.
i. Under section 57, any offence under the statute may be compounded; while under section 61, the minister may exempt any person or class of persons from any provision of the statute.

Before going further, it is important to note that the Explanatory Statement and Preamble of POFMA explicitly state the various purposes of the legislation. What is of relevance here are the stated aims of preventing the electronic communication of false statements of fact and information manipulation, as well as the enabling of measures to counteract the effects of such communication.[9] The Ministry of Law also made the following clarifications: the law targets falsehoods and not opinions, criticisms, satires, and parodies; the primary remedial measure would be corrections, and not removal or take-down of content; criminal offences would apply only to malicious actors seeking to undermine society; and any decision by the government over what is false can be overridden by the courts on appeal (Ministry of Law 2019a). These positions were reiterated in Parliament during the debate over the bill (Singapore Parliament 2019b), as well as in various cases.[10]

What then are some of the issues that arise from the provisions that have been set out above? Here, I identify several issues that have been raised by various parties; these also correspond to the chief concerns that were raised in Parliament during the debate over the legislation (A. Lim 2019b). The first and perhaps most obvious one is that concerning definitions, the chief example of which is what constitutes a 'false statement of fact' under section 7. Section 7 does not define the phrase,[11] but the definitional clause (section 2(2))

does clarify that the determination of a statement of fact is based on a 'reasonable person' analysis, and that false statements would include partly or wholly misleading statements.[12] Jurisprudence has clarified that both analyses of whether a statement is a statement of fact and whether that statement is false involve objective assessments.[13] Notwithstanding this attempt at disambiguation, it is not always easy to distinguish between facts and opinions, characterisations, and misimpressions that result from decontextualisation.[14]

Uncertainty in the aforementioned determination can be attributed to the inherent nature of language, and is not always one that can be resolved through legislation or judicial reasoning. The Singapore Courts have grappled with the issue of whether a statement constitutes fact or opinion. For instance, in the 2022 High Court decision of *Singapore Democratic Party v Attorney-General*, the statement in question read: 'the HDB chief executive [Dr Cheong] said that Singapore's population density would increase from 11,000 people per sq km to 13,700 people per sq km between now and 2030. Given our land area, this means that our population would go up to nearly 10 million by 2030'.[15] The appellant sought to argue that the deductive process by which the numerical figure of 10 million was arrived at was an opinion as opposed to fact.[16] This argument was made based on case law stipulating that a comment, as opposed to a statement of fact, is something that can be reasonably inferred to be a deduction.[17] Nonetheless, the court found that the opening statement of the article in question, '[t]he idea of Singapore increasing its population to 10 million did not originate from the SDP' sufficed to show that the statement in question purported to be a report of what Dr Cheong had said in his statement.[18] However, this seems to leave unaddressed whether such a deduction, in isolation, could constitute opinion.

This difficulty can be further examined by gleaning at the Ministry of Law's paper, titled 'How the Protection from Online Falsehoods and Manipulation Act applies' (Ministry of Law 2019b). The paper outlines, *inter alia*, the difference between statement of fact and opinion, providing examples to the effect. For example, an opinion would be where F, a professor of medicine, states that a certain treatment is effective, citing a clinical study to support his statement. However, if the study F cites does not stand for what F says it stands for, F's statement would be a false statement of fact. While in theory, this seems to accord with logic, in practice, F still has to interpret what the clinical study may stand for. Thus, notwithstanding efforts to disambiguate the uncertainties in the distinction between fact and opinion, certain greyness remains.

The illustrations we have cited highlight the importance of needing POFMA to strike the right balance between allowing internet discourse on matters of public interest to flourish (maybe accepting that mistakes on the facts necessarily have to be made before a consensus can be reached) and ensuring that misunderstandings that lead to hate speech and possibly violence are minimised. This brings us to the second issue, which concerns the allocation of powers in the fact-determination process. POFMA leaves no doubt that it is the government (minister) who determines whether a false statement of fact has been made, and whether it is in the public interest to order correction or stop directions, general or targeted corrections, or disabling of access of end-users. But when it comes to online hate speech regulation, are there superior alternatives?

The usual rejoinder to letting the government decide is to let either the intermediary or the online community self-regulate, though often it is a combination of both. For instance,

Facebook and Twitter have in place user term agreements and community standards, reporting and flagging mechanisms, moderators, geographical reach limiters, algorithmic filters, and inhouse legal teams to track and curtail potential hate speech – indeed in some jurisdictions, they seldom wait for a government or court order before acting (to remove the post, suspend the user, and so forth).[19]

The problem though – apart from a lack of mandate, since it is the government that should be facilitating and protecting speech – is the corporations' abject lack of transparency whenever intermediaries moderate content or suspend or ban accounts, leading some to conclude that intermediaries have the proclivity to censor based on their own political ideologies – Facebook and Twitter have routinely been accused of considering contentious social or political issues as not up for because of the supposed potential for generating hate speech, but there are neither appeal mechanisms nor reasons given when they limit or block such speech (Guynn 2018). Seen in this light, putting aside the difficulty of determining what is hate speech, maybe it is actually more democratic and accountable for the government, rather than intermediaries, to make the call on when the freedom of expression should be limited. I should add briefly that having the courts make the call is not a viable alternative, for reasons that the government have highlighted during the debate: courts are not designed to act quickly, both in terms of getting a hearing date and allowing the full presentation of evidence in adversarial system of law (Singapore Parliament 2019b).

But even if one accepts that the government is best placed to make the call on whether something qualifies as hate speech, something might be said about the obstacles to challenging the government's decision. Once the minister decides that directions, corrections, or disablement are appropriate, these orders then must be complied with until an application to vary or cancel them succeeds, or potentially heavy sanctions could follow (this is elaborated in the next issue). Further, while the minister's decision to refuse variation or cancellation is appealable to the courts, this necessarily involves litigation, which entails time, resources, and costs (Omar 2019) (though it was clarified that the appeal process will be simple (no court fees for the first three days) (Singapore Courts 2022), and fast (as little as nine days for the court to hear the case)) (Tham 2019d).

One could of course rationalise this approach on the basis that hate speech is irredeemably bad for society, and so it stands to reason to make it hard for the maker of the speech or the intermediary hosting the speech to challenge the government's decision; further, as much as the government is responsible for upholding rights, it also has the responsibility to protect competing rights, such as the safety and security of vulnerable groups. Parliament has recently highlighted POFMA to be relevant in this regard as it serves as a tool against the exacerbation of hate speech through falsehoods (Singapore Parliament 2021). Yet this presupposes that – apart from its ease of determination – such speech has no inherent value whatsoever, and the best remedy is governmental intervention. In the final analysis, once it is accepted that the government is better to place to determine the legality of content, there would invariably be transactional costs involved to challenge it. How these costs can be meaningfully managed may well go beyond the issue of costs of litigation in this specific context, but the relatively high costs of litigation in Singapore in general.

The last issue identified for present purposes is the proportionality of the consequences once the minister has decided that there is hate speech. Directions to stop

and orders to disable access are not extraordinary and are indeed logical responses once one accepts that hate speech warrants legislative and executive action, but corrections might be tricky to implement. For individuals, the maker of the statement is expected to declare that his statement is false and point to where the truth is to be found on terms to be decided by the government. For intermediaries, by virtue of them being passive entities even though they are often passive entities that do not generate (but sometimes can control) user content, they have to communicate corrections on terms to be decided by the government. Because it is the government that decides the extent of corrections required, for the reasons stated above, this might be preferable to letting the intermediaries decide.

Having said that, in the context of hate speech, it seems unlikely in most cases that the government would opt for corrections – stop or disablement orders would more likely be preferred.[20] One would also imagine that in Singapore, hate speech, especially that pertaining to race or religion, is a presumptively egregious offence, given historical sensitivity to racial and religious disharmony within the community (Public Service Division 2015).[21] A person found guilty of such speech could be fined up to $50,000 and imprisoned for up to 5 years;[22] if he refuses to abide by any government order to stop or modify his statement, he could be subject to further sanctions in the form of a fine up to $20,000 and imprisonment of up to 12 months.[23] The fining of an intermediary of up to $1 million may not seem much if we have in mind the revenues of social media giants such as Facebook and Twitter, but this amount is for each breach and not a global cap.[24]

All things considered, it seems there is, at multiple junctures, a perceptible potential for chilling speech and self-censorship, but as earlier stated, the purpose of this comment is to compare POFMA with international norms, or at least those that are identifiable. This give us a better idea as to whether POFMA under- or over- regulates what has clearly become a societal menace, and also provides some yardsticks on what the best practices could be in combatting online hate speech – after all, the validity of any criticism of POFMA is best seen in the light of international human rights law and practices, for any other point of reference would not have the same pedigree of international consensus. Additionally, the view that effective Internet governance requires intergovernmental collaboration in formulating harmonised rules is increasingly gaining traction.[25] It is perhaps for these reasons that the Ministry of Law was minded to include an annex on international developments when introducing POFMA (Ministry of Law 2019a), as was the Select Committee in its report (Singapore Parliament 2018). With all that said, how does POFMA hold up under international legal scrutiny?

3. The international human rights framework for regulating online hate speech

Any discourse on international human rights, especially where speech is concerned, must invariably begin with the International Covenant on Civil and Political Rights (ICCPR) – the preeminent multilateral human rights treaty[26] that has been ratified by more than 170 of the 190-plus UN member states (more will be said soon about Singapore not being one). What must be observed at the outset is that although the freedom of expression is one of the fundamental rights protected by the covenant, this freedom[27] is not absolute, and can be limited by states.[28] Article 19(3) of the ICCPR permits the restriction of the freedom of

expression on the grounds of respect of the rights or reputations of others, or the protection of national security, public order, public health or morals.[29] The same provision states, however, that states in interfering with this right must show that any such restriction is provided by law and is necessary; while the threshold for prescription by law is low and gives states some latitude to frame laws broadly, the element of necessity requires the pursuit of a legitimate aim, the existence of a pressing social need, and a proportionate response that has no less restrictive alternatives (what is sometimes referred to as the three-part test).[30]

Then there is article 20(2), which unlike article 19 that requires the state to protect the freedom of expression, instead requires the state to prohibit certain kinds of speech. Specifically, it obligates states to prohibit speech that advocates national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence.[31] This has been referred to by the UN High Commissioner for Human Rights and various UN Special Rapporteurs as the hate speech prohibition, and has been confirmed to apply to online speech as well (UN Human Rights Council 2013, 29). It bears reiteration, therefore, that the duty of states with respect to speech is twofold: to not impede unduly, but also to ensure it does not cause harm. To be even more precise, states are permitted to regulate speech even if the speech does not meet the criteria of article 20(2) – they can regulate speech on the grounds in article 19(3), provided that the requirements of legality and necessity are met.

In determining whether there is hate speech, factors to be considered include the context in which the speech was made, the status of the maker of the speech, the intent of the maker, the content and form of the speech, the reach of the speech, and the likelihood and imminence of harm (inchoate acts are therefore included) (UN Human Rights Council 2013, 29). Although it has been said that any restrictions should only be imposed on 'the most severe and deeply felt form of opprobrium' (UN Human Rights Council 2013, 29) and must also fulfil the aforementioned prescription and necessity requirements (UN Human Rights Council 2018a, 8), it is widely accepted that the ICCPR conception of the limits on freedom of expression differ from, say, the traditional US conception, which sets the high watermark for permissible speech.[32] Recent UN bodies have even claimed that the approach towards online hate speech should be 'zero tolerance', in unequivocal opposition to any ideal of relying heavily on the marketplace of ideas to counteract hate speech.[33]

Operating in the backdrop is the doctrine of margin of appreciation or margin of discretion, which claims that because national authorities are best placed to balance conflicting fundamental human rights based on each state's unique social context and circumstances, because there is no one-size-fits-all solution when assessing the justifiability of interferences with human rights. However, there are obstacles to concluding that this doctrine enjoys unqualified universal support. For instance, while the European Court of Human Rights (ECtHR) has consistently affirmed it over decades of its jurisprudence,[34] the UN body which monitors the implementation of the ICCPR has been more tentative in its embracement.[35] This position by the Human Rights Council has also reiterated by the former UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye (UN General Assembly 2019, 27). From this standpoint, any invocation of the margin of appreciation outside the context of the European Convention on Human Rights[36] should, in principle at least, be done with some

caution.[37] Instead, absent customary international law to the contrary, the text of the ICCPR should remain the first port of call when determining the scope of the freedom of expression, the interpretation of which is of course guided by the rules of treaty interpretation set out in the Vienna Convention on the Law of Treaties.[38]

The above establishes the contours for regulating freedom of expression in general. Things become considerably trickier when internet websites enter the framework. For a start, the ICCPR was originally conceived to regulate affairs between the state and its people rather than the state and media entities, let alone intermediaries such as social media websites, the chief purveyors of virulent content (UN Human Rights Council 2018a); further, before the advent of the Internet, the freedom of expression was largely exercised on platforms fundamentally different from the most popular types of media today in terms of barriers to entry, costs, editorial intervention, speed of dissemination, reach, the potential for virality, privacy and anonymity, and the likelihood of fabrication and decontextualisation. The net result is that in the context of speech made on the Internet, it becomes unclear who should be responsible for content regulation such as the curbing of hate speech – is it still the state, as was the case before online freedom of expression was possible, or do internet websites, and in particular social media companies, now bear the main responsibility? And even if the allocation of responsibility can be determined, as alluded to above, what exactly should be done to remedy the breach, and what consequences should follow if the breach is not remedied?

In this regard, we can look at how online hate speech and intermediaries have been regulated in various parts of the world. It used to be the case that the importance of self- regulation, as well as the need to avoid broad and sweeping restrictions on internet content, were acknowledged almost universally – by the UN Special Rapporteur on Freedom of Opinion and Expression, the Organisation for Security and Co-operation in Europe, the Organisation for American States, the African Commission on Human and Peoples' Rights, to name but a few regional and international bodies (Organisation for Security and Co-operation in Europe 2011).

It also used to be the case that there was near-universal recognition that because the Internet has become the principal means by which people exercise their freedom of expression, attempts to restrict online speech on matters of public interest should be condemned (La Rue 2011, 58). Because of this, the default position in many jurisdictions was that even though they often had billions in revenue, the responsibilities imposed on intermediaries were either unarticulated or minimal, and speech (including hate speech) could only be lawfully restricted upon the issuance of a court order or executive order from the authorities (La Rue 2011, 43; Marino 2013, 138).

Complementing this was the intermediaries' repeated assurances that less restrictive alternatives of speech regulation worked more effectively to curtail hate speech – apart from the aforementioned user term agreements and community standards, reporting and flagging mechanisms, moderators, and algorithmic filters to put them on notice of problematic content, in more extreme cases intermediaries could also limit access, suspend accounts, block accounts, demonetise accounts, remove posts, or reduce the visibility of trending posts (MacKinnon et al. 2014). Indeed, apart from copyright infringement and child sexual abuse and exploitation material – for which both can be detected with success rates with the right software – there was no expectation that intermediaries had to do much with respect to content regulation (Keller 2018).

However, giving intermediaries a free pass is no longer the state of affairs. The ECtHR, for instance, now applies various factors for the determination of intermediary liability when an intermediary is perceived to be ineffectual and inefficient in policing content amounting to hate speech.[39] While the court recognises that value judgments, opinions, potentially defamatory speech, and crude expressions with a low register of style are still permissible, where an intermediary allows for content amounting to 'manifest expressions of hatred' to remain on its platform, it would be held accountable for being asleep at the wheel, especially if it has the means to determine the legality of posts.[40] In part, this was a rebuke to the notion that the marketplace of ideas would work well enough on the Internet, especially since intermediaries are often guilty of perpetuating echo chambers by controlling the content that users interact with based on the users' political preferences – any previously held assumptions about counter-narratives neutralising bad speech were thus shattered (Samples 2019).

At first, intermediaries resisted taking on this responsibility in monitoring content. The claim was that there was too much content on social media to monitor, and it was unreasonable to impose any duty to regulate when they had neither creative nor editorial oversight for information that could be published by anyone at anytime and anywhere – not to mention the perennial issue of the difficulty in determining hate speech in different contexts. Facebook, for instance, continually insisted for years that it was ill-placed to tackle the problem of fake news facilitating hate speech being shared on its platform before grudgingly accepting that it had to implement measures to fact- check news articles – only to be accused of introducing political bias into such checks (Vanian 2018; Zuckerberg 2016); its introduction of the Oversight Board has been met with a fair bit of criticism as well (Paul 2021). For Twitter, it was only in late-2018 that it revised its hate speech policy (Matsakis 2018), while YouTube continues to be criticised for not being proactive enough in taking down videos that encourage or depict racial and religious violence.[41] But increasing waves of national legislation have followed the ECtHR's lead, even enhancing the burden that is to be placed on intermediaries.

The most notable example is probably Germany, which has grappled with immigrant and refugee assimilation in recent years. It passed the Network Enforcement Act in 2017. Under that law, absent any specific agreement with law enforcement authorities, social network intermediaries are to remove 'manifestly unlawful' content within 24 h of receiving a user complaint, as opposed to executive or judicial notice; this time period is only extended to 7 days for unlawful content that is not 'manifestly unlawful' (Bundesministerium der Justiz 2017). Within this framework, intermediaries are expected to consult their own lawyers in determining whether something would count as illegal hate speech and the swiftness of the takedown required – failure to comply with the law may lead to fines of up to 50 million Euros (Spiegel 2018), a far cry from the hundreds of Euros the ECtHR is used to sanctioning for online hate speech.[42] It came as little surprise that the Act was criticised by the UN Human Rights Committee for its potential impact on the freedom of expression. The UNHRC pointed out that 'responsibility for the removal of such content is assigned to social media companies and not subject to judicial oversight, thereby limiting access to redress in cases where the nature of content is disputed' (UN Human Rights Committee 2021). Further amendments to the Act to expand the scope of responsibility of social network intermediaries have been met with strong opposition from intermediaries (Noyan 2022).

Even France[43] and the UK,[44] long the bastions of laïcité and the 'right to offend' respectively, have moved towards adopting something similar. So too in countries outside Europe, such as India.[45] And in the wake of the Christchurch shootings in neighbouring New Zealand, Australia passed amendments to its Criminal Code, under which intermediaries face penalties of up to 10% of the platform's annual turnover if they do not remove violent material expeditiously (Attorney-General for Australia, et al. 2019). Indeed, so-called rights-conscious jurisdictions now have, or will soon have laws that are more severe as compared to jurisdictions such as Malaysia (where the maximum fine is around USD 20,000)[46] and Russia (where the maximum fine is around USD 15,000) (Griffiths and Hodge 2019).

Lest it be assumed that the situation is somehow different for liability for individuals – it is not. It is no longer uncommon for jurisdictions to provide for fines and imprisonment sentences to punish those that knowingly spread fake news and disinformation that are tantamount to hate speech. States such as Bangladesh (Woollacott 2018), Egypt (Magdy 2019), and Malaysia (CLJ Law 2018), which mete out both considerably hefty fines and imprisonment sentences to tackle the so-called spread of disinformation that compromises national stability and security, are no longer anomalous. In addition to Germany (DW 2016) and the UK (Gov.UK 2020) also plans to imprison persons guilty of online disinformation campaigns, following the lead by the likes of Italy (Horton 2018) and Kenya (Gathright 2018). These states clearly do not think intermediary responsibility diminishes individual responsibility.

4. Where Singapore stands and other recommendations to improve the regulatory framework

If there was once a time that the exercise of freedom of expression online was to be freely celebrated for giving everyone a potentially powerful voice, that time has, in my view at least, passed. Whereas previous forms of ground-breaking mediums – be it the microphone, radio, broadcast television, or the Internet – were eventually found not to have fundamentally altered the landscape for freedom of expression such as to warrant paradigm shifts in regulation, states now consider social media to be the genuine game-changer (for the worse). This is reflected most obviously in how the interpretation of the ICCPR has changed drastically in the form of the new and emerging online hate speech laws we have just surveyed above; the Internet and social media may be ubiquitous, but it has not resulted in a unified understanding of how best to regulate the problem both in terms of offence creation and the punishments that follow. Whereas the previous emphasis was on invoking article 19(3) of the ICCPR only in extreme scenarios, states have now in effect pivoted to using article 20(2) as the default starting point, complemented by the use of the margin of appreciation and contiguous international human rights obligations such as those found in ICERD. The Rabat Plan used for determining hate speech is also decidedly open-ended enough without drawing any bright lines – much will depend on the facts of each case and the circumstances of each jurisdiction, a situation no doubt complicated by the amplifying nature of social media. The composite picture that emerges from all of this is not one that is terribly reassuring for those looking to discern discrete rules of international law in this domain – and one has not even begun to discuss if a clear line can be drawn between incitement laws

and anti-blasphemy laws, the latter of which has been condemned by the UN (Shaheed 2017) but can, in certain situations where statements about race or religion are made, be hardly distinguishable from the former.

As far as Singapore is concerned, it has not ratified the ICCPR, and so it is not bound by it as a matter of treaty law.[47] But given the very high ratification numbers,[48] putting aside issues of enforcement, there is some force in the claim that articles 19 and 20 of the ICCPR have attained the status of customary international law.[49] Yet even assuming the two articles bind Singapore, the aforementioned three-part test for justifying interferences with article 19 can probably be surmounted with some ease. Specifically, vague or broad as some of POFMA provisions may be, the threshold for prescription for law has always been low and almost never successfully argued before international tribunals, and in any event, to the extent that the prescription requirement can be satisfied with safeguards,[50] the minister's decision can be judicially reviewed and would constitute such safeguards. On the next requirement, few would, or can, quibble with the suppression of online hate speech as fulfilling a pressing social need and constituting a legitimate aim, notwithstanding the aforesaid difficulties in identifying what might qualify as hate speech.[51] But, also as mentioned, the Rabat Plan looks at multiple factors such that what may be hate speech in one place may not be hate speech in another – the factors of likelihood of harm, for instance, is going to be so context-sensitive. When viewed through an international lens that often minimises the review of domestic measures, this is not going to be an insurmountable hurdle. Lastly, the proportionality of POFMA sanctions, be it for individuals or intermediaries, would likely be situated in the middle of the spectrum of punishments for online hate speech when compared to current and developing international norms across the board.[52] And if the margin of appreciation does apply (at any given limb of the three-part test), this will only bolster Singapore's position, considering its political history with respect to managing racial and religious relations.

Notwithstanding the above, apart from the issues previously highlighted and analysed, the regulatory framework can be improved in the following ways in the light of what has been developing internationally. After all, given that international developments are now veering more towards the Singapore-style of governance for regulating online content, this is a rare opportunity for us to lead the way in setting standards. First, the current tenor of POFMA may give the impression that, unlike the case for individual offenders who may already face punishment for posting hate speech online, intermediaries are better off waiting for the government to issue an order before acting to remove hate speech hosted on their platforms instead of adopting a more proactive stance in monitoring hate speech. While it has been argued – as has the UN recently (UN Human Rights Council 2018a) – that the government is better placed than intermediaries to determine what is online hate speech and that a government order is probably the most democratic option when it comes to this type of censorship (UN Human Rights Council 2016, 51–55), this should not leave the tools that intermediaries already have (filters, complaint mechanisms, moderators to enforce community standards, and so forth) without teeth. There would be situations in which intermediaries would, through their own mechanisms or otherwise, be put on notice about potentially problematic content – for instance, in the wake of a terrorist attack, or on the eve of elections. When this happens, intermediaries should work in concert with the authorities to decide on the best course of

action, rather than revert to private censorship or do nothing at all. Preferably, however, any decision made this way should have some element of public access so that a hybrid form of private-public censorship does not ensue. If we look for example at Facebook's recent mass-bans of American and British public figures, their decisions have not been based on facts but simply their own feelings and political preferences (Paul 2019).[53] Similarly, Twitter's Musk-era content moderation policies recently drew fire from the European Union for its 'arbitrary approach to reinstating banned users'.

The second point is a related one: transparency in any decision-making remains important for public confidence in the system to be maintained. Indeed, a commentator has suggested that POFMA should 'require a Minister's order to identify the relevant falsehood, set out what the true position is, identify the specific public interest involved and how it is threatened by the falsehood, and articulate why the order is both proportionate and necessary' and also 'expressly require any order to be proportionate to the nature of the falsehood and the degree of harm to the public interest' (Singh 2019, 9 and 13). The same commentator has also suggested that as a matter of practice, the government 'should provide an annual summary of the Ministerial orders issued, the facts and circumstances of each case, the reasons for the specific Ministerial orders, the number of appeals … and the outcome of the appeals … This annual review would enable Parliament to decide if the law is properly achieving its stated legislative aims' (Singh 2019, 21). Adopting these two suggestions does incur some costs and is not strictly required even by international standards, but one should broadly be in agreement with the suggestions insofar as they promote transparency and Singapore continues to pride itself on being one of the most transparent countries in the world (A. Lim 2019a).

I do not think, however, as some commentators have suggested, that an additional layer of checks should exist in the form of an independent body (Tham 2019c).[54] Expediency is key to the efficient (and cost-effective) operation of POFMA, and a check already exists in the form of the courts. What is of greater importance is the accountability aspect. While the Ministry of Law has stated that the government's reasons for each order will be guided by subsidiary legislation, it may be more prudent to build this into the primary legislation. Subsidiary legislation takes time to develop and does not have the same binding effect as primary legislation, and as pointed out by several Nominated Members of Parliament during the debate, can be more easily amended (in a negative sense) than primary legislation (Singapore Parliament 2019b). Given the uncertainties over what would constitute online hate speech, requiring reasons to be given from the very first orders that are issued pursuant to POFMA would provide important certainty and strengthen the rule of law. This certainly would help not just with regard to what constitutes an offence, but also why certain orders were made and considered proportionate (which is separate from the proportionality of the sanctions that flow from violating the orders).

Next, I foresee difficulties in justifying the inclusion of messaging services as intermediaries that would fall under POFMA (whether in relation to online hate speech or otherwise). Under section 2, one of the examples given for what counts as an intermediary is 'internet-based messaging services'; section 3 also confirms that material is considered to be communicated for the purposes of POFMA so long as it is sent through the Internet. However, messaging services are meant to facilitate private communications. In contrast, the whole point of posting on social media is to hope that the content goes viral – which is hard to achieve unless there is some publicity of the content, even if it is only being

circulated within closed groups. When queried in Parliament about the problems raised by end-to-end encryption methods utilised by such closed platforms, the Senior Minister of State for Law, Mr Edwin Tong Chun Fai, identified that 'additional ways of dealing with the harm' would be utilised, such as through the issuance of a General Correction Order (Singapore Parliament 2019b). Another difference between a messaging service and social media is that of reach. Even before recent changes to limit the forwarding of messages (Iyengar 2019), apps like WhatsApp and Telegram just do not have the same design architecture and features as Facebook or Twitter to facilitate sharing of messages with the same scale.

More importantly, there is also the issue of privacy. In pretty much any given case, how is one supposed to monitor what is being shared between users of messaging services without some sort of surveillance?[55] This could explain why in countries where messaging services have been regulated for online hate speech, the governmental response has seldom deviated from a complete shutdown of those services.[56] When this was raised in Parliament, it was not stated how the government would be alerted of problematic content being shared through such services. Instead, the Senior Minister of State said that 'in closed spaces, people are more susceptible to emotive falsehoods, because these are the spaces inhabited by the familiar and the trusted.'[57] Of course, one of the ways to work around this is to have the messaging service prevent the sharing of certain websites or certain online articles. But this reintroduces the problems of private censorship, which on balance, is probably not desirable, even if the censorship is limited to reducing visibility and is based on user feedback.[58] Ultimately, in deciding to include messaging services, one should bear in mind that the pursuit of security still needs to be tempered by some sense of proportionality and efficacy. This challenge may well be made more acute as 'deep fakes' become more prevalent and difficult to verify, but that is something we would leave to for a future endeavour.

However, it must be noted that ever since it was passed into law, POFMA has yet to be deployed in regulating online hate speech. As of June 2022, it has been invoked 38 times since it came into force in 2019, with more than half of these instances targeted at correcting COVID-19 misinformation (POFMA Office 2022).[59] With respect to online hate speech, they have been punished under the Penal Code instead,[60] and in light of the Singapore government's recent announcement to de-criminalise sex between two consenting adult males, it is anticipated that new laws may be needed to regulate dialogue between groups that hold opposing views on this subject (Zalizan 2022). There may, therefore, be questions about the suitability and necessity of POFMA as a tool for regulating against online hate speech. While POFMA only targets the communication of false statements of fact, the scope of hate speech extends well beyond statements of facts (UN Human Rights Council 2013, 29). Existing laws focused on the intent and effect of speech[61] rather than the content of speech may provide a more direct, albeit less efficient, method of regulation. The key distinguishing factor of POFMA may lie in the mechanisms available to mitigate the effects of online hate speech.

At the same time, one cannot disregard the broad scope of falsehoods that the POFMA targets – extending far beyond the regulation of hate speech. To this end, the justifiability and effectiveness of the POFMA mechanisms in its restriction of the freedom of expression remains to be in contention. Having regard to these questions, the overall significance and efficacy of POFMA remains to be seen.

None of the above should detract from the fact that POFMA may, in some sense, be seen as the forerunner in paradigms for speech regulation. Although Singapore has typically been characterised as less liberal in permitted speech, by allowing posts to remain online so long as the authorities have a right to rebut it is more liberal than what has been seen elsewhere. For instance, consider cancel culture in American universities (The Economist 2022). Furthermore, POFMA places the responsibility for the removal of online hate speech with the government. This stands in contrast to the aforementioned approaches adopted in other jurisdictions, which often involve the controversial imposition of duties on social media providers to remove harmful content.

Of course, it is necessary to recognise the limitations on the exportability of POFMA model. One must consider the unique contextual circumstances of Singapore, which may support the introduction of legislation like POFMA. Singaporeans have high levels of trust in the government,[62] which in turn, may have contributed to the receptiveness of POFMA (Ho 2019). In other jurisdictions, where trust in the government is lower, social media providers may be preferred as an 'independent' arbiter in content regulation.

Finally, the introduction of legislation like POFMA does not address problems arising from regulation of online speech by social media providers. It is undeniable that regulation of online speech occurs regardless of whether there is domestic legislation. The mere fact that POFMA creates a publicly accountable process through which social media providers may be compelled to remove content, does not prevent social media providers from regulating online content on their platforms. As earlier discussed, such regulation may raise issues owing to lack of transparency.

5. Conclusion

This article has sought to situate POFMA within the larger international movement towards combatting hate speech. While one ought to recognise POFMA's limitations with respect to its operation and wide scope, it would be a hasty move to dismiss the mechanisms utilised in POFMA in considering the possible regulatory approaches. This is especially as countries actively strive towards determining the most appropriate and effective ways of combatting online hate speech.

Notes

1. The government has also stated that efforts in improving media and information literacy and promoting fact-checking would be taken as well. Indeed, simply developing an appetite for longform debates would go some way in neutralising simplified soundbites on, say, power structures and identity politics that seed the ground for online hate speech. In November 2022, the Singapore government also proposed to introduce a bill that would require intermediaries to take more proactive steps in removing 'harmful content'. A discussion of that bill will have to be a subsequent endeavour.
2. The result was a 317-page report.
3. This is defined in section 3(2) as a statement that is made available to one or more end-users in Singapore on or through the Internet, MMS, or SMS.
4. *Cf* section 3(1)(*e*) of the Sedition Act 1948. Note, however, that the Sedition Act 1948 has been repealed: see Sedition (Repeal) Act 2021.
5. *Cf* section 8(1)(*a*) of the Maintenance of Religious Harmony Act 1990.

6. In recent times, the link between the two has manifested in its most violent form in debates on immigration policy and refugee policy (UN Human Rights Council 2018b).
7. Commentators have added that '[POFMA] does not provide a deadline within which the minister must decide. This should be expressly set out in subsidiary legislation (as is usually the case), and should ideally be kept short so as to enable the aggrieved person to have recourse to the courts without undue delay.' (Omar 2019). The Ministry of Law later said it would consider this suggestion (Shanmugam 2019). This eventually translated to regulation 14 of the Protection from Online Falsehoods and Manipulation Regulations 2019.
8. See also the government powers under the Broadcasting Act 1994.
9. It has also been suggested that POFMA can be applied in a preventive way that focuses on the intent to do harm, without harm necessarily materialising (Tham 2019b).
10. Mainly, the issue of whether a statement is a fact or an opinion for the purposes of setting aside a correction direction or stop communication direction under section 17(5)(*b*) of POFMA has been repeatedly raised and discussed: see for instance *Singapore Democratic Party v Attorney-General* [2020] SGHC 25 at [24]–[32]; *The Online Citizen Pte Ltd v Attorney-General* [2020] SGHC 36 at [46]–[51]; *The Online Citizen Pte Ltd v Attorney-General* [2022] SGHC 177 at [35]–[39]. Moreover, the notion that the courts remain the arbiter of truth and falsity, ie, that the Minister's assessment that a false statement of fact has been made is subject to a final determination by the court as to whether or not it is correct (*The Online Citizen Pte Ltd v Attorney-General* [2021] 2 SLR 1358 at [60]–[61]).
11. Indeed, the ambiguities present in POFMA extends even to terms that have been defined – 'in the public interest' being the prominent example – the definition seems circular when we compare the language between sections 4 and 10 (or 11, 12, 20, 21, 22, and 23 for the matter).
12. Though in a different context, article 16 of the General Data Protection Regulation (2019) Directive 95/46/EC has shown that the concept of incomplete accounts of information is difficult to ascertain and adjudicate upon.
13. *The Online Citizen Pte Ltd v Attorney-General* [2021] 2 SLR 1358 at [158]–[159].
14. See for instance *Salov v Ukraine* App No 65518/01 (ECtHR, 6 September 2005) at [113].
15. *Singapore Democratic Party v Attorney-General* [2022] SGHC 100 at [55].
16. *Singapore Democratic Party v Attorney-General* [2022] SGHC 100 at [56].
17. *Chen Cheng and another v Central Christian Church* [1998] 3 SLR(R) 236 at [34].
18. *Singapore Democratic Party v Attorney-General* [2022] SGHC 100 at [59]–[60]. While this case went on appeal, this finding was not disturbed by the Court of Appeal: see *Singapore Democratic Party v Attorney-General* [2022] SGCA 56 at [57]–[58].
19. This will be elaborated in the next section.
20. Even then, this potentially places intermediaries that operate in multiple jurisdictions between a rock and a hard place – while POFMA requires them to comply with directions as there is no defence of duty under law, other jurisdictions may sanction them for unduly restricting the freedom of expression of the person who posted the content.
21. See section 3(1)(*e*) of the Sedition Act; section 8(1)(*a*) of the Maintenance of Religious Harmony Act 1990.
22. Section 7(1)(*a*) of POFMA.
23. Section 15(1)(*a*) of POFMA.
24. Section 42(1)(*b*) of POFMA.
25. This is evinced by the UN' Internet Governance Forum in 2018.
26. The International Bill of Rights comprises the ICCPR, the International Covenant on Economic, Social and Cultural Rights, 993 UNTS 3 (entry into force 3 January 1976) and the Universal Declaration of Human Rights, 10 December 1948, 217A(III).
27. Article 19(2) states that everyone 'shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.'
28. *Cf* article 19(1), which states that everyone 'shall have the right to hold opinions without interference.'

29. Article 18(3), which pertains to the contiguous right of freedom of religion, similarly restricts religious expressions that might cause harm to public safety, order, health, morals, or the fundamental rights and freedoms of others. See also principle 6 of The Johannesburg Principles on National Security, Freedom of Expression and Access to Information (1995).
30. This has been the unanimous position across all international human rights bodies, be it the UN (see for instance *Corinna Horvath v Australia* UN Doc CCPR/C/110/D/1885/2009 (HRC, 27 March 2014) at [3.11]), the Strasbourg court (see for instance *Avram v Moldova* App no 41588/05 (ECtHR, 5 July 2011) at [24]), the Inter- American court (see for instance *Tristán Donoso v Panama*, Preliminary Objection, Merits, Reparations, and Costs Judgment (IACtHR, 27 January 2009) at [56]), or the African Human Rights court (see for instance *Interights v Mauritania* AHRLR 87 Comm no 242/2001 at [78]–[79]). See also UN Human Rights Committee, 'General Comment 34' (12 September 2011) UN Doc CCPR/C/GC/34 at [22]–[34].
31. See also article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination, 660 UNTS 195 (entry into force 4 January 1969).
32. This has been the case since the SCOTUS decision interpreting the Amendment I of the Constitution of the US in *Brandenburg v Ohio*, 395 US 444 (1969).
33. For instance, consider the Third Committee's Experts recommendation that states must develop zero-tolerance policies towards white supremacy – and other extremist ideologies, hate speech and incitement to hatred (UN 2018).
34. See for instance *Bédat v Switzerland* App no 56925/08 (ECtHR, 29 March 2016) at [54]; *Bierski v Poland* App no 46342/19) (ECtHR, 20 October 2022) at [42] and [49]; *Pavlov v Russia* App no 31612/09 (ECtHR, 11 October 2022) at [75].
35. See for instance *Ilmari Länsman v Finland* UN Doc CCPR/C/52/D/511/1992 (HRC, 14 October 1993) at [9.4].
36. ETS 5 (4 November 1950). Having said that, there is no material difference between article 19 of the ICCPR and its equivalent in the ECHR in terms of the grounds in which the freedom of expression may be limited.
37. The European cases that have consistently upheld that speech that offends and shocks the conscience is permissible (see for instance *Sürek v Turkey* App no 23927/94 (ECtHR, 8 July 1999) at [58]). This should, likewise, be treated with some caution.
38. 1155 UNTS 331 (entry into force 27 January 1980). Typical canons of interpretation are set out in article 31.
39. See for instance *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v Hungary* App no 22947/13 (ECtHR, 2 May 2016) at [69]; *Drousiotis v Cyprus* App no 42315/15 (ECtHR, 5 October 2022) at [41]. The factors include the context of the comments, the measures applied by the intermediary company to prevent or remove defamatory comments, the liability of the actual authors of the comments as an alternative to the intermediary's liability, and the consequences of the domestic proceedings for the applicant company. For the (similar) position under the European Court of Justice, see *Google France, Google Inc v Louis Vuitton Malletier* SA C-236/08 (CJEU, 23 March 2010) at [120].
40. *Delfi AS v Estonia* App no 64569/09 (ECtHR, 16 June 2015) at [142]–[156]. In that case, the contents in question were left online for no less than six weeks.
41. Active monitoring is, of course, only one step away from pre-emptive censorship and prior restraint.
42. See also *Delfi AS v Estonia* App no 40287/98 (ECtHR, 16 June 2015) at [27]–[31].
43. The French 'Loi Avia', a law aiming to regulate online hate speech, entered into force on 1 July 2020. However, the French Constitutional Court had earlier struck down the core provision of the law, which mandated the removal of hate speech from online platforms within 24 h of notice, finding this to be a breach of the freedom of expression (Boring 2020).
44. Following the publication of the Online Harms White Paper in April 2019, the UK introduced the Draft Online Safety Bill in May 2021: https://bills.parliament.uk/publications/49376/documents/2822
45. India introduced the Guidelines for Intermediaries and Digital Media Ethics Code Rules in February 2021, subjecting social media intermediaries to an expanded set of due diligence

obligations: https://www.meity.gov.in/writereaddata/files/Intermediary_Guidelines_and_Digital_Media_Ethics_Code_Rules-2021.pdf.

46. While Malaysia's Anti-Fake News Act of 2018 was repealed in late 2019, Malaysia adopted similar provisions in Emergency (Essential Powers) (No 2) Ordinance 2021, which aimed to tackle misinformation related to the pandemic (Schuldt 2021).
47. Singapore also subscribes to the dualist approach to the domestic incorporation of international law (Chen 2011).
48. The 19 states that have neither signed nor ratified do not include any major states.
49. The requisite elements are that of widespread and representative state practice and *opinio juris*: *North Sea Continental Shelf Cases*, ICJ Reports 1969 at [73]–[77].
50. *Liu v Russia (no 2)* App no 29157/09 (ECtHR, 26 July 2011) at [88].
51. See generally *Robert Faurisson v France* UN Doc CCPR/C/58/D/550/1993 (HRC, 2 January 1993).
52. Further, the assessment of whether the requirements of prescription by law, the existence of a pressing social need, and a proportionate response with no less restrictive measures must necessarily be conducted against the backdrop of the specific circumstances of a state and take into account the 'specific need on which they are predicated' (UN Rights Committee 2011).
53. A common tactic they use is to brand groups they disagree with as 'extremist' or 'far-right'. As private corporations, they can claim to be not obligated to uphold constitutional freedoms, but they should not be given free rein to do as they please either.
54. It was also stated in Parliament that such a body would not have the same democratic mandate as the government to make decisions (M. Z. Lim 2019).
55. As to whether there are privacy rights in Singapore, see Chen Siyuan, 'The Regulatory Framework for Aerial Imaging by Recreational Users of 'Drones' in Singapore: Old and Emerging Issues and Some Possible Solutions' (2017) 29(1) Singapore Academy of Law Journal 126.
56. See for instance Business Insider, 'The Turkish Government Reportedly Blocked WhatsApp and Other Social Media Sites' (4 November 2016): https://www.businessinsider.com/social-media-and-messaging-sites-blocked- in-turkey-2016-11/?IR = T; The Straits Times, 'Indonesian Government Lifts Ban on Telegram' (12 August 2017): https://www.straitstimes.com/asia/se-asia/indonesian-government-lifts-ban-on-telegram; CNA, 'Sri Lanka Social Media Shutdown Raises Fears on Free Expression' (23 April 2019): https://www.channelnewsasia.com/news/asia/sri-lanka-social-media-shutdown-raises-fears-on-free-expression- 11469136.
57. The Straits Times, 'Closed, Encrypted Communications Also Covered by Law' (8 May 2019): https://www.straitstimes.com/politics/closed-encrypted-communications-also-covered-by-law-edwin-tong.
58. See also The Straits Times, 'Facebook Rolls Out Fact-Checking Service in Singapore to Combat Fake News' (3 May 2019): https://www.straitstimes.com/tech/facebook-rolls-out-fact-checking-service-in-singapore-to- combat-fake-news; The Daily Wire, 'Journalism Schools Poynter Publishes List of 'Unreliable News'' (2 May 2019): https://www.dailywire.com/news/46703/journalism-school-poynter-publishes-list-ashe-schow.
59. As of 1 December 2021, POFMA had been invoked 33 times, with 19 targeted at COVID-19 misinformation (Chee 2021).
60. For instance, a man was charged over making Instagram Stories posts with deliberate intent to wound religious feelings of any person (Yeo 2021). Another man was also charged for making remarks on his Facebook and Instagram accounts to wound the feelings of the Christian community on several occasions (Kamil 2022).
61. Section 298 of the Penal Code 1871 targets persons who act with the intention of wounding the religious or racial feelings of any person.
62. The 2021Edelman Trust Barometer reflected that the trust index for trust in the government in Singapore was 76, well above the global average of 53 (Edelman 2021).

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

AG for Australia, Minister for Industrial Relations. 2019. "Tough New laws to Protect Australians from Live-Streaming of Violent Crimes". *Australian Government*. https://webarchive.nla.gov.au/awa/20210325132612/https://www.attorneygeneral.gov.au/media/media-releases/tough-new-laws-protect-australians-live-streaming-violent-crimes-4-april-2019.

Boring, Nicholas. 2020. "France: Constitutional Court Strikes Down Key Provisions of Bill on Hate Speech." *Library of Congress*. June 29. https://www.loc.gov/item/global-legal-monitor/2020-06-29/france-constitutional-court-strikes-down-key-provisions-of-bill-on-hate-speech/#:~:text=The%20Conseil%20constitutionnel%20agreed%20that,harmful%20to%20freedom%20of%20expression.

Bundesministerium der Justiz. 2017. "Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act)." July 12. https://www.bmj.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf?__blob=publicationF%20ile&v=2.

Chee, Kenny. 2021. "Singapore's Fake News Law Used 33 times to Date." *The Straits Times*. December 1. https://www.straitstimes.com/tech/tech-news/singapores-fake-news-law-used-33-times-to-date-including-19-times-against-covid-19.

Chen, Siyuan. 2011. "The Relationship between International Law and Domestic Law." *Singapore Academy of Law* 23: 350–366.

Chia, Chen Wei. 2018. "Sketching the Margins of a Borderless World." *Singapore Academy of Law* 30: 833–870.

CLJ Law. 2018. "Anti-Fake News Bill 2018." https://www.cljlaw.com/files/bills/pdf/2018/MY_FS_BIL_2018_06.pdf.

Council of Europe Committee of Ministers. 1997. "Recommendation No. R (97) 20 Of the Committee of Ministers to Member States on "Hate Speech" (Adopted by the Committee of Ministers on 30 October 1997 at the 607th meeting of the Ministers' Deputies)." October 30. https://rm.coe.int/1680505d5b.

DW. 2016. "Facebook User Jailed for Hate Comments." October 17. https://www.dw.com/en/german-court-sentences-facebook-user-to-jail-for-xenophobic-comments/a-36069082.

The Economist. 2022. "Pushback at Cancel Culture is Leading to New Educational Initiatives." February 26. https://www-economist-com.libproxy.smu.edu.sg/united-states/2022/02/26/pushback-at-cancel-culture-is-leading-to-new-educational-initiatives/.

Edelman. 2021. "Edelmen Trust Barometer 2021 – *Country Report* Trust in Singapore." https://www.edelman.com/sites/g/files/aatuss191/files/2021-03/2021%20Edelman%20Trust%20Barometer_Singapore%20Country%20Report_1.pdf.

Fox, Chris. 2019. "Websites to be Fined over Online Harms under New Proposals. *BBC*. April 8. https://www.bbc.com/news/technology-47826946.

Gathright, Jenny. 2018. "Kenya's Crackdown on Fake News Raises Questions About Press Freedom." *NPR*. May 19. https://www.npr.org/sections/thetwo-way/2018/05/19/612649393/kenyas-crackdown-on-fake-news-raises-questions-about-press-freedom.

Gov.UK. 2020. "Online Harms White Paper." December 15. https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper.

Griffiths, James, and Nathan Hodge. 2019. "New Law lets Russia Jail People who Disrespect the Government Online." *CNN*. March 7. https://edition.cnn.com/2019/03/07/europe/russia-internet-law-intl/index.html.

Guynn, Jessica. 2018. "Is Facebook too Liberal?" *USA Today*. https://www.usatoday.com/story/tech/news/2018/05/03/facebook-pledges-investigate-charges-bias-against-conservatives/574505002/.

Ho, Kim. 2019. "Half of Singaporeans Support Anti-Fake news Bill." *YouGov*. May 22. https://sg.yougov.com/en-sg/news/2019/05/22/half-singaporeans-support-anti-fake-news-bill/.

Horton, Alex. 2018. "A Man was Sentenced to 9 Months in Prison. His crime? Posting fake reviews on TripAdvisor." *The Washington Post*. September 12. https://www.washingtonpost.com/world/2018/09/12/man-tried-sell-fake-tripadvisor-reviews-hes-going-prison-after-landmark-ruling-italy/.

Human Rights Watch. 2021. "Singapore: 'Fake News' Law Curtails Speech." January 13. https://www.hrw.org/news/2021/01/13/singapore-fake-news-law-curtails-speech.
Iyengar, Rishi. 2019. "WhatsApp Tightens Limit on the Number of People you can Share Messages with." *CNN*. January 21. https://edition.cnn.com/2019/01/21/tech/whatsapp-forwarding-limits-india/index.html.
Kamil, Asyraf. 2022. "Man, 31, to be Charged with Intending to Wound Religious Feelings of Christians in Social Media Posts." *Today*. January 18. https://www.todayonline.com/singapore/man-31-be-charged-intending-wound-religious-feelings-christians-social-media-posts-1792676.
Kaye, David. 2018. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*. Geneva: UN Human Rights Council.
Kaye, David. 2019. "Mandate of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression." https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/Legislation/OL_SGP_3_2019.pdf.
Keller, Daphne. 2018. "Internet Platforms." *Hoover Institution's Aegis Paper Series*, no. 1807.
La Rue, Frank. 2011. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*. Geneva: UN Human Rights Council.
Lever, Rob. 2019. "Sri Lanka Social Media Shutdown Raises Fears on Free Expression." *Yahoo*. April 23. https://news.yahoo.com/sri-lanka-social-media-shutdown-raises-fears-free-172140320.html.
Lim, Min Zhang. 2019. "Why Independent Council to Review Govt's Actions is not Necessary." *The Straits Times*. 2019. https://www.straitstimes.com/politics/why-independent-council-to-review-govts-actions-is-not-necessary.
Lim, Adrian. 2019a. "Singapore Rises to Third Place in Annual Ranking of Least Corrupt Countries." *The Straits Times*. January 29. https://www.straitstimes.com/politics/singapore-rises-to-third-place-in-annual-ranking-of-least-corrupt-countries-global-study.
Lim, Adrian. 2019b. "Parliament: Shanmugam Addresses 5 Concerns over Proposed Fake News Law." *The Straits Times*. May 8. https://www.straitstimes.com/politics/parliament-law-minister-shanmugam-addresses-five-concerns-over-proposed-fake-news-law.
MacKinnon, et al. 2014. *Fostering Freedom Online*. UN: UN Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000231162.
Magdy, Samy. 2019. "Egypt Tightens Restrictions on Media, Social Networks." *AP*. March 20. https://apnews.com/article/1540f1133267485db356db1e58db985b.
Marino, Catalina Botero. 2013. *Freedom of Expression and the Internet*. Washington, DC: Inter-American Commission on Human Rights.
Matsakis, Louise. 2018. "Twitter Releases New Policy on Dehumanizing Speech." *Wired*. September 25. https://www.wired.com/story/twitter-dehumanizing-speech-policy/.
Ministry of Home Affairs. 2019. "Ministerial Statement on Restricting Hate Speech to Maintain Racial and Religious Harmony in Singapore, Speech by Mr K Shanmugam Minister for Home Affairs and Minister for Law." April 1. https://www.mha.gov.sg/mediaroom/parliamentary/ministerial-statement-on-restricting-hate-speech-to-maintain-racial-and-religious-harmony-in-singapore-speech-by-mr-k-shanmugam-minister-for-home-affairs-and-minister-for-law/.
Ministry of Law. 2019a. "New Bill to Protect Society from Online Falsehoods and Malicious Actors." April 1. https://www.mlaw.gov.sg/news/press-releases/new-bill-to-protect-society-from-online-falsehoods-and-malicious-actors#:~:text=Key%20Features%20of%20the%20Bill,the%20scope%20of%20the%20Bill.
Ministry of Law. 2019b. "How the Protection from Online Falsehoods and Manipulation Act Applies." https://www.academia.sg/wp-content/uploads/2019/07/How_POFMA_Applies.pdf.
Noyan, Oliver. 2022. "Big Tech Opposes Germany's Enhanced Hate Speech Law." *Euractiv*. 3 February. https://www.euractiv.com/section/internet-governance/news/german-reinforcement-of-hate-speech-law-faces-opposition-from-big-online-platforms/.
Omar, Siraj. 2019. "Protection from Online Falsehoods and Manipulation Bill." *The Straits Times*. May 1. https://www.straitstimes.com/opinion/a-more-calibrated-approach.
Organisation for Security and Co-operation in Europe. 2011. "Joint Declaration on Freedom of Expression and the Internet." June 1. https://www.osce.org/fom/78309.

Paul, Katie. 2019. "Facebook Bans Alex Jones, Other Extremist Figures. Reuters." May 3. https://www.reuters.com/article/us-facebook-extremists-usa/facebook-bans-alex-jones-other-extremist-figures-0idUSKCN1S82D7.

Paul, Kari. 2021. "Facebook Ruling on Trump Renews Criticism of Oversight Board." *The Guardian*. May 5. https://www.theguardian.com/technology/2021/may/05/facebook-oversight-board-donald-trump.

POFMA Office. 2022. "POFMA Action Taken Up to 30 June 2022." September 6. https://www.pofmaoffice.gov.sg/files/Tabulation-of-POFMA-cases-and-actions.pdf.

Posetti, et al. 2018. *Journalism, Fake News & Disinformation*. UN: UN Educational, Scientific and Cultural Organisation. https://en.unesco.org/sites/default/files/journalism_fake_news_disinformation_print_friendly_0.pdf.

Public Service Division. 2015. "Cultivating a Harmonious Society, Becoming One People." https://www.psd.gov.sg/HOPS/our-institutions/cultivating-a-harmonious-society-becoming-one-people/.

Reporters Without Borders. 2020. "RSF Denounces Singapore's Disregard of Press Freedom ahead of its Universal Periodic Review." October 19. https://rsf.org/en/rsf-s-denounces-singapore-s-disregard-press-freedom-ahead-its-universal-periodic-review.

Samples, John. 2019. "Why the Government Should Not Regulate Content Moderation of Social Media." *Cato Institute*. April 9. https://www.cato.org/policy-analysis/why-government-should-not-regulate-content-moderation-social-media#the-presumption-against-public-regulation-of-social-media.

Saudi Gazette. 2018. "5-Year Jail, SR3m Fine for Social Media Material that Disrupt Public Order." May 9. https://saudigazette.com.sa/article/542723/SAUDI-ARABIA/5-year-jail-SR3m-fine-for-social-%20media-material-that-disrupt-public-order.

Schuldt, Lasse. 2021. "The Rebirth of Malaysia's Fake News Law." April 13. *Verfassungblog on Matters Constitutional*. https://verfassungsblog.de/malaysia-fake-news/.

Shaheed, Ahmed. 2017. *Elimination of all Forms of Religious Intolerance (Interim Report of the Special Rapporteur on Freedom of Religion or Belief)*. August 28. UN: Human Rights Council.

Shanmugam, Kasiviswanathan. 2019. "NMPs Agree on Major Points of Falsehoods Bill." *The Straits Times*. May 3. https://www.straitstimes.com/opinion/nmps-agree-on-major-points-of-falsehoods-bill?utm_medium=Social&utm_campaign=STFB&utm_source=Facebook&fbclid=IwAR1G3HRKuS8R4j3bYg_CiKOWXRx-AvLBwPhj1iABgBSTZEPTV7cARZmkw04#Echobox=1556887416.

Singapore Courts. 2022. "Appeals under the Protection from Online Falsehoods and Manipulation Act (POFMA) (from 1 April 2022)." April 1. https://www.judiciary.gov.sg/civil/appeals-under-the-protection-from-online-falsehoods-and-manipulation-act-(pofma)-(from-1-april-2022).

Singapore Parliament. 2018. "Report of the Select Committee on Deliberate Falsehoods." September 20. https://sprs.parl.gov.sg/selectcommittee/searchPage?committee=Other%20Select%20Committee%20Reports&from=20-09-2018&to=20-09-2018.

Singapore Parliament. 2019a. "Restricting Hate Speech to Maintain Racial and Religious Harmony in Singapore." April 1. https://sprs.parl.gov.sg/search/#/sprs3topic?reportid=ministerial-statement-1170.

Singapore Parliament. 2019b. "Second Reading of the Protection from Online Falsehoods and Manipulation Bill." May 8. https://sprs.parl.gov.sg/search/sprs3topic?reportid=bill-366.

Singapore Parliament. 2021. "Efforts to Prevent Hateful or Inciteful Speech or Action Being Communicated or Distributed." April 10. https://sprs.parl.gov.sg/search/sprs3topic?reportid=written-answer-8934.

Singh, Harpreet. 2019. "Strengthening the Online Falsehoods Bill. *Singapore Law Watch*." April. https://www.singaporelawwatch.sg/Portals/0/190402%20Online%20Falsehoods%20Bill.pdf?fbclid=IwAR3.

Spiegel, Johanna. 2018. "Germany's Network Enforcement Act and its Impact on Social Networks." *Lexology*. August 6. https://www.lexology.com/library/detail.aspx?g=fb107efe-70ae-4e97-9913-5035aeeb518a.

Tan, Daron. 2021. "Singapore: ICJ calls on Government to Repeal or Substantially Amend POFMA." *International Commission of Jurists*. October 4. https://www.icj.org/singapore-icj-calls-on-government-to-repeal-or-substantially-amend-the-pofma-two-years-after-it-entered-into-force/.
Tham, Yuen-C. 2019a. "Parliament: Law against Online Falsehoods will not Stifle Free Speech." *The Straits Times*. April 1. https://www.straitstimes.com/politics/parliament-law-against-online-falsehoods-will-not-stifle-speech-ministers.
Tham, Yuen-C. 2019b. "Legislation Essential to Curbing Spread of Fake News. "*The Straits Times*. April 26. https://www.straitstimes.com/politics/pm-legislation-essential-to-curbing-spread-of-fake-news.
Tham, Yuen-C. 2019c. "Shanmugam Responds to Key Issues on Fake News Bill." *The Straits Times*. May 4. https://www.straitstimes.com/politics/law-minister-responds-to-key-issues-on-fake-news-bill.
Tham, Yuen-C. 2019d. "Appeal Process under Fake News Law will be Simple, Fast." *The Straits Times*. May 8. https://www.straitstimes.com/singapore/appeal-process-under-fake-news-law-will-be-simple-fast-shanmugam.
UN. 2018. "Global Cooperation and Regulation Key in Addressing Multilayered Threats Posed by New Technology." November 14. https://www.un.org/sustainabledevelopment/blog/2018/11/global-cooperation-and-regulation-key-in-addressing-multilayered-threats-posed-by-new-technology/.
UN General Assembly. 2019. "Promotion and Protection of the Right to Freedom of Opinion and Expression." October 9. https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/A_74_486.pdf.
UN Human Rights Committee. 2011. "General Comment No. 34." September 12. https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf.
UN Human Rights Committee. 2021. "Concluding Observations on the Seventh Periodic Report of Germany." November 30. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/357/46/PDF/G2135746.pdf?OpenElement.
UN Human Rights Council. 2013. "Rabat Plan of Action." January 11. https://www.ohchr.org/sites/default/files/Rabat_draft_outcome.pdf.
UN Human Rights Council. 2016. "Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression." May 11. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G16/095/12/PDF/G1609512.pdf.
UN Human Rights Council. 2018a. "Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression." April 6. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf.
UN Human Rights Council. 2018b. "Report of the Independent International Fact-finding Mission on Myanmar." September 12. https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_64.pdf.
Vanian, Jonathan. 2018. "Facebook CEO Mark Zuckerberg Admits Huge Mistake But Will Not Step Down." *Fortune*. April 5. https://fortune.com/2018/04/04/facebook-mark-zuckerberg-data-cambridge-analytica/.
Woollacott, Emma. 2018. "Bangladeshi Digital Security Act Draws Fire From EU." *Forbes*. September 28. https://www.forbes.com/sites/emmawoollacott/2018/09/28/bangladeshi-digital-security-act-draws-fire-from-eu/?sh=322e46160277.
Yeo, Shu Hui. 2021. "Man to be Charged over Making Religiously Offensive Social Media Posts." *The Straits Times*. December 19. https://www.straitstimes.com/singapore/courts-crime/man-to-be-charged-over-making-religiously-offensive-social-media-posts.
Zalizan, Taufiq. 2022. "Govt could Legislate against Cancel Culture if Right Solutions can be Found." *Today*. September 19. https://www.todayonline.com/singapore/legislate-cancel-culture-solutions-shanmugam-1993676.
Zuckerberg, Mark. 2016. "Status Update – 13 November 2016." *Facebook*. November 13. https://www.facebook.com/zuck/posts/10103253901916271.