

Public Consultation Paper on Enhancing Online Safety

Submission by **Cultivate SG**

20 December 2024



Table of Contents

I.	Introduction	3
II.	“Cultural Lag” and a Needed Catch-up on Online Harms	3
	A. Damage of Online Harms to Society	4
	B. Spotlight on Online Sexual Harms	5
	C. Our Survey Findings on Cancel Culture and the Online Space	8
	D. Limitations of Existing Laws	11
III.	Comments on Online Harms Measures	12
	A. Scope of Online Harms	13
	1) Intimate Image Abuse	13
	2) Child Abuse Material	16
	3) Online Statements Instigating Disproportionate Harm (“OSIDH”)	16
	4) Hate Speech	20
	5) Recommendation: Other Harms to be Included	22
	6) Other Considerations: Vulnerable Persons	22
	B. Role of the Proposed Agency on Online Harms	24
	1) Publication of the Agency’s Orders	24
	2) Avenues for Appeal	25
	C. Improving Accountability in the Online Space	26
	1) Disclosure of User Information	27

2) Platforms' Collection of User Information	29
IV. Conclusion	32

I. Introduction

1. Cultivate SG is a non-profit organisation which wants to see families and our society thrive for generations. We call this ‘social sustainability’. This involves individual rights and responsibilities, stable marriages, strong families, a cultural climate that supports personal and family growth, and social harmony. We believe that culture – as the sum total of values, beliefs and practices of people in society – is not a battle to be fought, but a garden to be cultivated.
2. We had previously observed a “cultural lag” concerning online harms, where technological advancements have outpaced our legal and moral norms, and the existing legal protections are insufficient to address these harms.¹
3. Thus, we are supportive in general of the proposed new measures to enhance online safety, detailed in the Public Consultation Paper published by the Ministry of Law (“**MinLaw**”) and Ministry of Digital Development and Information (“**MDDI**”) on 22 November 2024.²

II. “Cultural Lag” and a Needed Catch-up on Online Harms

4. Cultivate shares the concerns of MinLaw and MDDI with regard to online harms and agrees on the need for legal reform.
5. In this section, we will address:
 - A. The damage of online harms to society;
 - B. A special focus on online sexual harms;
 - C. Our survey findings on cancel culture and the online space;
 - D. Limitations of existing laws.



¹ Darius Lee, “As technology outpaces law on online harms, new solutions are needed” *The Straits Times* (19 October 2023): <https://www.straitstimes.com/opinion/as-technology-outpaces-law-on-online-harms-new-solutions-are-needed>.

² Ministry of Law and Ministry of Digital Development and Information, “Public Consultation Paper on Enhancing Online Safety: Empowering Singaporeans to Seek Relief from Harmful Online Content and Conduct, and Hold Responsible Parties Accountable” (22 November 2024) (“**Consultation Paper**”).

A. Damage of Online Harms to Society

6. Singapore is highly digitally connected. A 2023 Infocomm Media Development Authority (“**IMDA**”) report revealed that 99% of resident households are connected to the internet and 98% of households with school-going children have access to computers.³ Such high levels of connectivity are a source of strength and vulnerability, as many aspects of daily life in Singapore (both public and private) take place online, and the lines between public and private life can often be blurred or breached.
7. As a communitarian society which emphasises “shared values”,⁴ a common set of norms – including in our interactions both online and offline – is fundamental to social harmony and the continued flourishing of our society. In the words of Sir Patrick Devlin, “without shared ideas on politics, morals and ethics no society can exist”.⁵
8. Each instance of online harm affects a victim in profound ways, causing mental, psychological, emotional and sometimes physical or financial harm. However, the consequences extend beyond the victim. If these become prevalent, they can collectively poison the tone of social discourse, damage the common good and harm society. Such societal harms include:
 - (a) **Damage to social harmony.** A toxic online space where harmful content proliferates is detrimental to social integration, as different groups may retreat into “virtual gated communities” or ideological “echo chambers” around similar interests or viewpoints.⁶ Some may choose to stay away entirely. This undermines genuine interaction between people in society and sows the seeds of societal polarisation and “us versus them” attitudes in the long run.
 - (b) **Harm to children and young persons.** Given their special vulnerabilities, children are at particular risk in the online space, if they are targeted by malicious actors to be sexually or ideologically groomed or exposed to inappropriate content. Some may also be manipulated or influenced by others (including their peers) to be involved in inappropriate or illegal behaviours. If toxic attitudes and behaviours are normalised, children and young persons may also become acculturated and influenced to participate in such behaviours and perpetrate online harms against others.

--

³ Infocomm Media Development Authority, “Singapore Digital Society Report 2023”: <https://www.imda.gov.sg/-/media/imda/files/infocomm-media-landscape/research-and-statistics/singapore-digital-society-report/singapore-digital-society-report-2023.pdf>.

⁴ Parliament, *White Paper on Shared Values* (Cmd. 1 of 1991).

⁵ Sir Patrick Devlin, “The Enforcement of Morals” in *Proceedings of the British Academy*, vol. 45 (1959), 129 at 137 to 138.

⁶ “How social media filter bubbles and algorithms influence the election” *The Guardian* (22 May 2017); Walter Quattrocchi, Antonio Scala and Cass R. Sunstein, “Echo Chambers on Facebook” (June 13, 2016), online: <https://ssrn.com/abstract=2795110>.

- (c) ***Detrimental to a family-friendly culture.*** The proliferation of sexually explicit or violent images online, harassment, racist and other harmful content, can be detrimental to the well-being of children, families and society. These can distort healthy understanding of human dignity, of the relationships between men and women, of sex and marriage, and of pro-social interactions in society.

B. Spotlight on Online Sexual Harms

9. One aspect of online harms that raises our deep concern is that of online sexual harms. With the unprecedented connectivity, blurring and obfuscation of lines between the public and private, as well as the capacity for online anonymity and pseudonymity, online sexual harms proliferate the internet. These harm not only individuals (including children) but also the wider moral ecosystem necessary for marriage, family and other healthy relationships to thrive.
10. Online sexual harms include but are not limited to:
- (a) ***Pornography.*** Pornography involves sexually explicit content that is primarily intended to cause sexual arousal in the viewer. It is degrading, dehumanising and corrupting, objectifying people as instruments of pleasure instead of persons with inherent dignity.⁷ It impairs and damages the ability of people to develop healthy meaningful and appropriate relationships. It is potentially addictive and hijacks novelty-seeking tendencies;⁸ this can lead individuals to seek out increasingly violent, extreme and illegal content.⁹ With the unprecedented accessibility, affordability and anonymity of the internet, the digital age has exacerbated the spread of pornography. As a result, it can enable, perpetrate or lead to various sexual crimes (see paragraph 12 below).
- (b) ***Deepfake sexual content.*** This involves situations where a person's likeness is imposed onto sexually explicit images using artificial intelligence. In a recent local incident, some Singapore Sports School students became victims of deepfake nude photos, which were created and spread by their peers.¹⁰
- (c) ***Child sexual abuse material ("CSAM").*** According to a 2023 report, the majority of CSAM offences investigated by the police since 2020 pertained to the possession,

--

⁷ Robert P. George and Shaykh Hamza Yusuf, "Pornography, Respect, and Responsibility: A Letter to the Hotel Industry" *The Public Discourse* (9 July 2012): <https://www.thepublicdiscourse.com/2012/07/5815/>.

⁸ Paula Blanca, et al., "Novelty, conditioning and attentional bias to sexual rewards" (2016) 72 *Journal of Psychiatric Research* 91.

⁹ Children's Commissioner (UK), "'A lot of it is actually just abuse': Young people and pornography" (January 2023): <https://assets.childrenscommissioner.gov.uk/wpuploads/2023/02/cc-a-lot-of-it-is-actually-just-abuse-young-people-and-pornography-updated.pdf>.

¹⁰ "Police investigating deepfake nude photos of Singapore Sports School students" *The Straits Times* (13 November 2024): <https://www.straitstimes.com/singapore/police-investigating-deepfake-nude-photos-of-singapore-sports-school-students>.

access and distribution of the same.¹¹ Beyond such cases, there is also so-called “self-generated” CSAM where children are groomed, deceived or extorted into producing and sharing a sexual image or video of themselves. Alarming, a 2023 report by the Internet Watch Foundation – a United Kingdom-based charity – found that **the majority of CSAM online (254,071 out of 275,652 webpages, or 92%) contained ‘self-generated’ imagery.** Children aged 11 to 13 years appeared most frequently.¹²

(d) **Sexual grooming.** Through the connection and anonymity afforded by social media, adults have a channel which they can use to sexually groom children and young people, including through false pretences and identities.¹³

11. The scale and age at which children are exposed to online sexual harms are highly concerning. A survey by international think-tank DQ Institute found that **more than 3 in 10 (16%) children aged 8 to 12 years of age have been involved in online sexual behaviours.** These behaviours include having searched or visited websites with sexual content, having proactively downloaded, sent or received online sexual content, and having had sexual conversations online with strangers.¹⁴ In the United Kingdom, a groundbreaking police analysis on reported child sexual abuse and exploitation found that 52% of such cases involved children offending against other children.¹⁵ Part of the reason, according to the police, was the accessibility of harmful and abusive pornography online. 25% of such cases involved online offences of indecent images of children.¹⁶
12. There are also companies and platforms that enable and profit off such harmful content (including illegal content). A telling 2020 *New York Times* exposé titled “The Children of Pornhub” revealed the shocking extent to which Pornhub was profiting off illegal content, including child sexual abuse, non-consensual intimate images, and



¹¹ “S’pore police probed 96 cases of child sex abuse material offences since new laws in 2020” *The Straits Times* (16 April 2023): <https://www.straitstimes.com/singapore/police-probed-96-cases-of-child-sex-abuse-material-offences-since-new-laws-were-introduced-in-2020>.

¹² Internet Watch Foundation, “‘Self-generated’ child sexual abuse”: <https://www.iwf.org.uk/annual-report-2023/trends-and-data/self-generated-child-sex-abuse/>. The notion of being “self-generated” is somewhat of a misnomer, since children lack sufficient maturity and agency and should not be held responsible for their own abuse.

¹³ See, for example, “‘I felt very violated’: S’porean youth opens up about brush with online sex predator at 14” *Mothership* (16 February 2020): <https://mothership.sg/2020/02/online-child-grooming/>.

¹⁴ DQ Institute, “2018 National DQ Impact Report”: https://www.singtel.com/content/dam/singtel/sustainability/Sustainability_reports_PDF%27s/2018_DQ_Singapore_National_Report.pdf.

¹⁵ National Police Chief’s Council Vulnerability Knowledge and Practice Programme (UK), “National Analysis of Police-Recorded Child Sexual Abuse & Exploitation (CSAE) Crimes Report”: <https://www.vkpp.org.uk/assets/Files/Publications/National-Analysis-of-police-recorded-CSAE-Crimes-Report-2022-external.pdf>.

¹⁶ National Police Chief’s Council Vulnerability Knowledge and Practice Programme (UK), “Child Sexual Abuse and Exploitation Analysis Launched”: <https://news.npcc.police.uk/releases/vkpp-launch-national-analysis-of-police-recorded-child-sexual-abuse-and-exploitation-csae-crimes-report-2022>.

violent content.¹⁷ In response, financial institutions Mastercard, Visa and Discover blocked customers from using their companies' credit cards to make purchases on Pornhub.¹⁸

13. Although everyone could potentially be a victim of online sexual harms, it is without a doubt that certain groups are more vulnerable than others to certain harms:
- (a) **Women and girls.** The Sunlight Alliance for Action found in a 2022 report that **females aged 25 to 34 are most likely to experience gender-based online harms**; compared to 72.1% of males, only 60.9% of women felt safe from online harms.¹⁹ Likewise, Cultivate's survey findings on social discourse found that women tend to be less comfortable than men in the online space.²⁰ These problems are exacerbated for girls and young women, as the anonymity afforded online has made it easier for predators to approach, communicate with and stalk them while avoiding accountability.²¹
 - (b) **Boys.** Boys are disproportionately exposed to online sexually explicit content. A **2016 survey by Touch Cyber Wellness found that 9 in 10 teenage boys in Singapore aged 13 to 15 years of age have watched or read sexually explicit materials within the past year**, whereas the figure was 8% for girls.²² Exposure to such content is damaging for children's development, rendering them less able to develop healthy attitudes, views and relationships with the opposite sex.²³
 - (c) **Lower-income groups.** Research from the Organisation for Economic Co-operation and Development ("OECD") found that socioeconomically advantaged students are more likely than disadvantaged students to search for information or read news online, whereas disadvantaged students are more likely to chat or play



¹⁷ Nicholas Kristof, "The Children of Pornhub" *New York Times* (4 December 2020): <https://www.nytimes.com/2020/12/04/opinion/sunday/pornhub-rape-trafficking.html>.

¹⁸ "Mastercard, Visa and Discover cut ties with Pornhub following allegations of child abuse" *CNN Business* (14 December 2020): <https://edition.cnn.com/2020/12/14/business/mastercard-visa-discover-pornhub/index.html>.

¹⁹ Sunlight Alliance for Action, "Tackling online harms, especially those targeted at women and girls" (August 2022): <https://www.mddi.gov.sg/files/Press%20Releases%202022/mci%20sunlight%20report%20fa-ed-compressed.pdf>.

²⁰ 52% of women reported being comfortable discussing controversial issues with people who do not share the same views in the online space, as opposed to 56% of men. (Cultivate SG, "Marriage, Family and Social Discourse Survey 2024" (15 November 2024): <https://cultivate.sg/research-recommend/marriage-family-and-social-discourse-survey-2024/>. We commissioned Toluna to conduct the survey of 2,000 Singapore Citizens and Permanent Residents.)

²¹ Jonathan Haidt, *The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness* (New York: Penguin Press, 2024) at 172.

²² "9 in in 10 teen boys in Singapore exposed to porn: Survey" *The Straits Times* (29 June 2016): <https://stomp.straitstimes.com/singapore-seen/singapore/9-in-in-10-teen-boys-in-singapore-exposed-to-porn-survey>.

²³ See, for example, the research in Pathmendra P, *et al.*, "Exposure to Pornography and Adolescent Sexual Behavior: Systematic Review" *Journal of Medical Internet Research* (2023) 25:e43116: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10015350/>.

video games online.²⁴ In light of this digital divide, it is foreseeable that experiences with the internet may exacerbate socioeconomic disadvantages as children in lower-income households may be at greater risk of harm.

C. Our Survey Findings on Cancel Culture and the Online Space

14. Another area of concern is that of cancel culture and the online space. Broadly speaking, cancel culture is the practice or tendency of engaging in mass withdrawal of support for someone as a way of expressing disapproval and exerting social pressure. It is linked to “call-out culture”, which involves the public criticism or faulting of someone.
15. Cancel culture is aggravated due to an interplay of a complex mix of factors involving human psychology and the capabilities of Internet and social media:
 - (a) ***Thinking fast and slow.*** Human beings tend to pass judgments based on intuition and emotion when making quick decisions, while tending to be more deliberative and logical in their judgments when given more time.²⁵
 - (b) ***Inclinations towards tribes and tribalism.*** “Morality binds and blinds.”²⁶ As social beings, humans are inclined to form tribes based on various identities – whether such identities are based on race, religion, political ideology, or various forms of social identity – and to favour their in-group over out-group. This may result in “us” versus “them” or “friend” versus “foe” tendencies.²⁷
 - (c) ***Characteristics of the internet and social media.*** The internet and social media have had aggravated the dark sides of human tendencies, due to the speed of communication, anonymity, filters and engagement-based ranking systems which create virtual echo chambers.²⁸ Content can be crafted to go “viral” by preying on human vulnerabilities, thereby raising the risk that extreme, polarising and divisive content is promoted.²⁹



²⁴ OECD, “Are there differences in how advantaged and disadvantaged students use the Internet?” *PISA in Focus* – 2015/07 (July).

²⁵ Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2013).

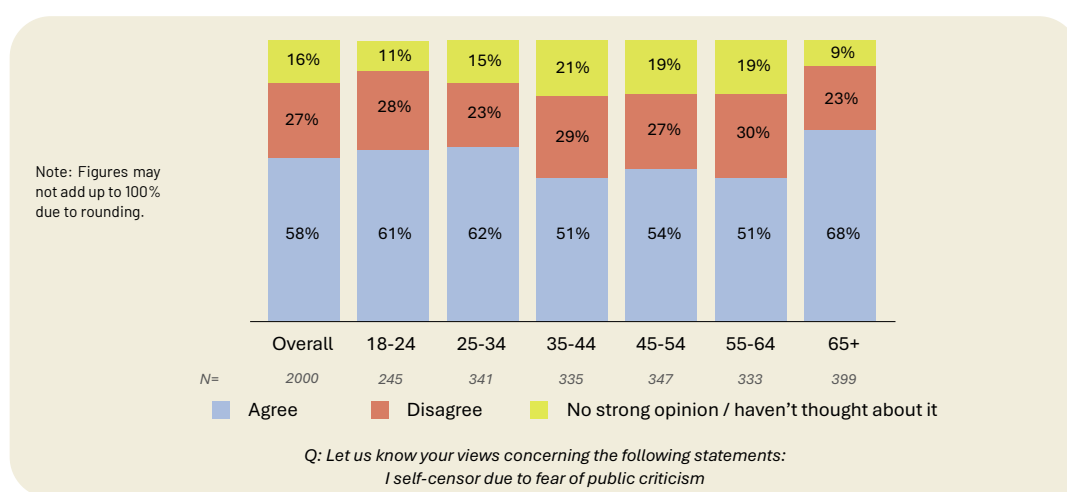
²⁶ See, generally, Jonathan Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (New York: Pantheon Books, 2012).

²⁷ Ministry of Finance, “Speech by Minister for Finance Lawrence Wong at IPS-RSIS Conference on New Tribalism and Identity Politics on 23 November 2021” (23 November 2021), online: <https://www.mof.gov.sg/news-publications/speeches/speech-by-minister-for-finance-lawrence-wong-at-ips-rsis-conference-on-new-tribalism-and-identity-politics-on-23-november-2021>.

²⁸ “How social media filter bubbles and algorithms influence the election” *The Guardian* (22 May 2017); Walter Quattrociocchi, Antonio Scala and Cass R. Sunstein, “Echo Chambers on Facebook” (June 13, 2016), online: <https://ssrn.com/abstract=2795110>.

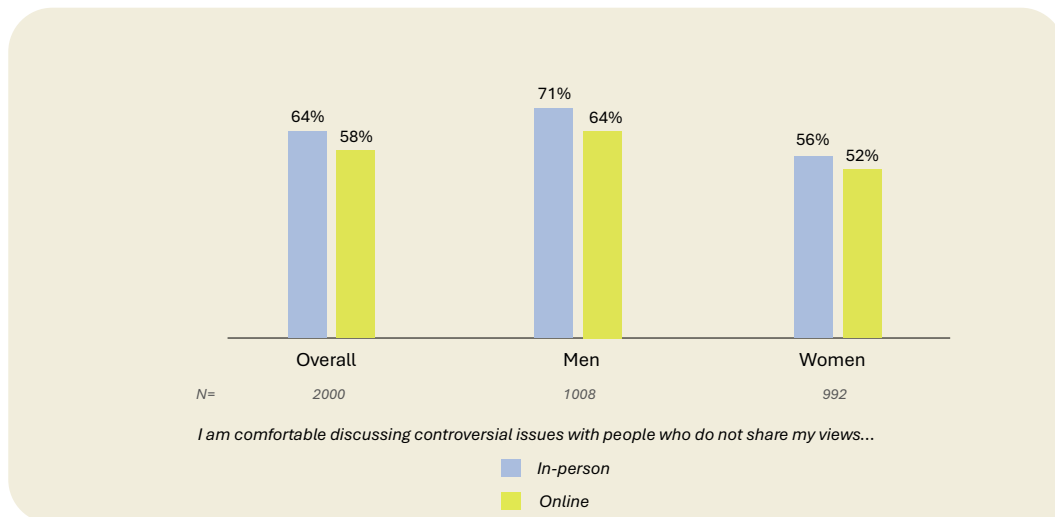
²⁹ See the criticism of Facebook Whistleblower Frances Haugen, in her interview by the Center for Humane Technology, “A Conversation with Facebook Whistleblower Frances Haugen” (18 October 2021), online: <https://www.humanetech.com/podcast/42-a-conversation-with-facebook-whistleblower-frances-haugen>.

16. This results in a toxic situation *where the speed of information flow exceeds the speed of rational thought*. Emotive, inflammatory and offensive content can easily spread like wildfire, whereas calm, rational and balanced perspectives tend to receive less attention. People who are targeted by online mobs are subject to immense amounts of pressure, and may be affected when employers, business partners and others sever their ties as a knee-jerk reaction.
17. Accordingly, the harm done by cancel culture is the *disproportionate penalty of social exclusion, including any related material loss (for real or alleged wrongdoing)*.
18. Cultivate’s recent survey³⁰ found significant tendencies towards self-censorship, where almost 6 in 10 respondents (58%) report that they self-censor due to the fear of public criticism.

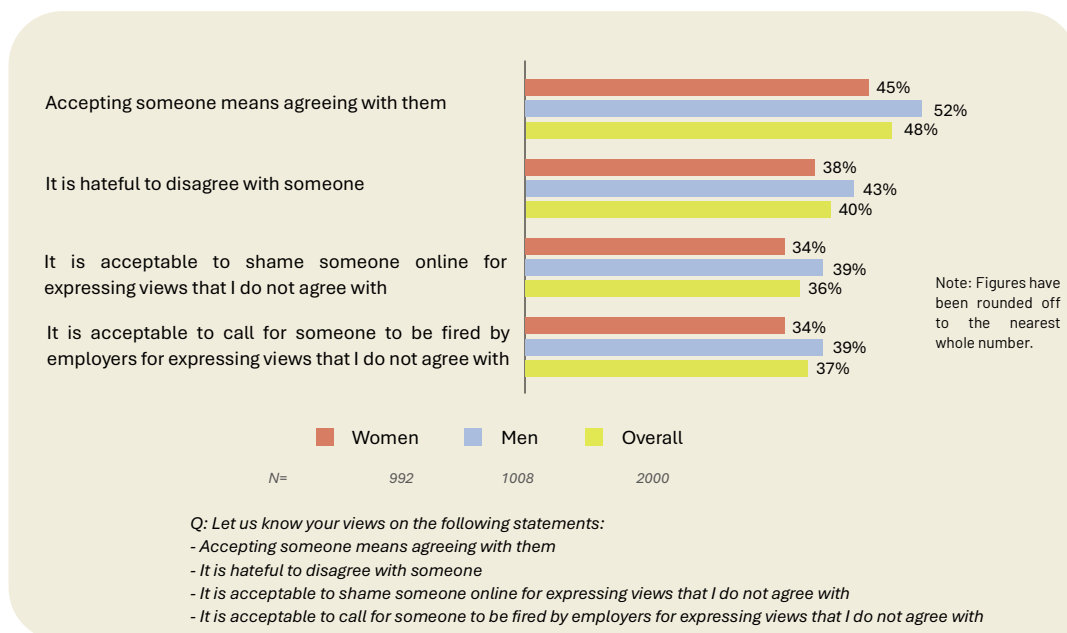


19. When discussing controversial issues with people who do not share the same views, respondents generally feel more comfortable doing so in-person (64%) rather than online (58%). Women are less comfortable than men discussing controversial issues with people who do not share their views in both settings.

³⁰ Cultivate SG, “Marriage, Family and Social Discourse Survey 2024” (15 November 2024): <https://cultivate.sg/research-recommend/marriage-family-and-social-discourse-survey-2024/>. We commissioned Toluna to conduct the survey of 2,000 Singapore Citizens and Permanent Residents.



20. The levels of self-censorship and relative discomfort in the online space may arise in part from certain attitudes of some people towards shaming and ‘cancelling’ those who do not share their views. More than 1 in 3 respondents say it is acceptable to shame someone online or call for someone to be fired by employers for expressing views that they disagree with. Men are more inclined towards such attitudes than women.



21. On the other hand, it is heartening that a majority of respondents in our survey still disapprove of shaming someone online or calling for someone to be fired by employers for expressing views that they disagree with.

22. They are also open to dialogue and make friends across controversial views, where almost 7 in 10 (67%) respondents say that they do not allow controversial views to stand in the way of them making friends. More than 7 in 10 (74%) respondents believe that it is possible for people with strongly opposing views to dialogue with each other.


23. Thus, there is a need to further encourage and foster such pro-social behaviours in the online space, while addressing negative and harmful online behaviours, so as to foster a better tone of social discourse.

D. Limitations of Existing Laws

24. We agree with the Government that existing laws are limited in their ability to address online harms, arising from limitations in various features of the law and the nature of the online space today.

25. Among the limitations of existing laws are:

- (a) ***The scope of existing laws and remedies.*** Harms such as intimate image abuse, impersonation, cancel campaigns and misuse of inauthentic material are not adequately covered in the civil law at present; victims of these harms may be limited in their options. While criminal laws may provide some remedy, the harms could overburden police resources if they are widespread, and it is understandable that police have to prioritise investigations of more severe crimes.
- (b) ***Anonymity and pseudonymity in the online world.*** Internet anonymity, pseudonymity and the use of Virtual Private Networks (“VPNs”) poses a well-known problem to existing legal frameworks, whether civil or criminal, since wrongdoing can be perpetrated anonymously. This limits the ability of potential claimants in pursuing the wrongdoers in civil cases, especially if there are multiple anonymous accounts involved.³¹
- (c) ***Speed and cost of the legal process.*** Due to the nature of the legal process, lawsuits are typically slower and costlier. Compared to the speed that online harms can be perpetrated and the damage they cause, the harm is often already long done by the time the process is concluded.
- (d) ***Limits of remedies under existing laws.*** Damages are hardly an adequate remedy for certain kinds of online harms, such as intimate image abuse. Furthermore, injunctions against publications of non-consensual intimate images may be reduced in effectiveness, if these images are circulated through multiple online accounts or websites.
- (e) ***Psychological and other practical limitations.*** Victims of online harms may experience severe levels of stress, and may not be in a psychological state to engage lawyers or pursue civil claims in court; taking such legal action may add to the stress instead. In addition, in cases such as intimate image abuse, a woman


³¹ In order to pierce the anonymity of the Internet and ascertain the proper party to be sued in a civil claim, a party may have to sue the service provider (such as Google), in order to compel the service provider to disclose the identity of the person behind the anonymous online posts. (See, for example, *Kabbabe v Google LLC* [2020] FCA 126)

may not want to relive the negative experiences or face the perpetrator, or various forms of stigma arising from the experience. Typically, a victim of online harms simply wants the harm to stop and go away.

26. For these reasons, Cultivate believes that there is currently a “cultural lag”, where technological advancements have outpaced our legal and moral norms,³² and agrees with the Government’s steps to strengthen protections in the online space.

III. Comments on Online Harms Measures

27. Constitutionally in Singapore, it is well recognised that freedom of speech and expression is not absolute but must be exercised responsibly and balanced against wider public interests, including the rights of others to be safe from harm.³³ In the final analysis, a carefully calibrated balance should be struck between the competing interests.
28. We note that the Government plans to introduce a complaints mechanism to provide victims with timely help when faced with online harms, under the purview of a government agency dedicated to supporting victims of online harms and enhance online safety (the “Agency”). In addition, it is proposed to introduce new statutory torts on online harms and to increase accountability through improved user information disclosure.³⁴
29. In this section, we will address:
 - A. Scope of online harms;
 - B. Role of the proposed Agency on online harms;
 - C. Improving accountability in the online space.


³² Darius Lee, “As technology outpaces law on online harms, new solutions are needed” *The Straits Times* (19 October 2023): <https://www.straitstimes.com/opinion/as-technology-outpaces-law-on-online-harms-new-solutions-are-needed>.

³³ *Wham Kwok Han Jolovan v Public Prosecutor* [2021] 1 SLR 476 at para. 20; *Chee Siok Chin and others v Minister for Home Affairs and another* [2006] 1 SLR(R) 582 at para. 42. “The tension between the individual’s right to speak and/or to assemble freely and the competing interests of security and/or public order calls into play a delicate balancing exercise involving several imponderables and factors such as societal values, pluralism, prevailing social and economic considerations as well as the common good of the community.” (*Chee Siok Chin and others v Minister for Home Affairs and another* [2006] 1 SLR(R) 582 at para. 52)

³⁴ Consultation Paper, para. 15.

A. Scope of Online Harms

30. We set out our comments on the scope of online harms as follows:

1) Intimate Image Abuse

31. The Consultation Paper proposes to include intimate image abuse, defined as: “Online communication of an intimate image or recording depicting or appearing to depict the victim, without the victim’s consent and which is likely to cause the victim harassment, alarm, distress or humiliation.”³⁵

32. We make the following comments:

(a) ***Definition of “intimate image or recording”***. Currently, Section 377BE(5) the Penal Code contains a definition of “intimate image or recording”. We recommend adopting the same definition.

(b) ***“Harassment, alarm, distress or humiliation”***. It should *not* be necessary for the victim to additionally prove or show that the communication of the intimate image or recording “is likely to cause the victim harassment, alarm, distress or humiliation”.³⁶ Such a requirement in law may unnecessarily raise the legal threshold for this kind of online harm since, for example, it is difficult for a victim to prove that an image posted in a members-only forum (to which the victim does not belong) is “likely” to cause the victim harassment, alarm, distress or humiliation.³⁷ We would recommend that this requirement be removed from the definition, and it should be sufficient for a victim to show that the intimate image or recording is being communicated without consent.³⁸



³⁵ Annex C, Consultation Paper.

³⁶ We note that the definition here is similar to some extent to Section 377BE(1) of the Penal Code regarding the distribution or threat to distribute intimate image or recording, where the element of “humiliation, alarm or distress” has also been included. Because of the scope of the offence addresses “distribution” or “threat to distribute” and attracts criminal penalties, the element of “humiliation, alarm or distress” may arguably be more relevant in that context. However, we submit that the position is different with regard to the present proposal on online harms.

³⁷ This is one of the limitations of existing laws, as we had raised in our article “As technology outpaces law on online harms, new solutions are needed” *The Straits Times* (19 October 2023): <https://www.straitstimes.com/opinion/as-technology-outpaces-law-on-online-harms-new-solutions-are-needed>.

³⁸ For example, the Australian position does not require proof of “harassment, alarm, distress or humiliation”:

- Under Section 16 of the Online Safety Act, it is sufficient that the person did not consent, in order to qualify as a “non-consensual intimate image of the person”. Certain exemptions are contained in the statute.
- Under Section 91Q(1) of the New South Wales Crimes Act, the offence of distributing an intimate image without consent is defined as follows: “A person who intentionally distributes an intimate image of another person (a) without the consent of the person, and (b) knowing the person did not consent to the distribution or being reckless as to whether the person consented to the distribution, is guilty of an offence.”

(c) **Exceptions or non-application.** There may be circumstances where it may not be appropriate to invoke the Agency’s powers to address the alleged online harm. This may include a situation of a so-called “copyright claim”, for example, when a pornographic image of a person posted on one website is circulated on another website without the consent of the person. It may be possible to exclude such cases by a robust analysis of the scope of consent, or by carving out exceptions or non-applications from the definition of “intimate image abuse”.

33. We would invite the relevant Ministry to consider adopting or adapting the definition of “non-consensual intimate image” under Australia’s Online Safety Act (the founding Act of Australia’s e-Safety):

Extracts from Australia’s Online Safety Act

16 Non-consensual intimate image of a person

For the purposes of this Act, if:

(a) an intimate image of a person is provided on:

- (i) a social media service; or
- (ii) a relevant electronic service; or
- (iii) a designated internet service; and

(b) the person did not consent to the provision of the intimate image on the service; and

(c) the provision of the intimate image on the service is not an exempt provision of the intimate image;

the intimate image is a non-consensual intimate image of the person.

86 Exempt provision of an intimate image

(1) For the purposes of this Act, if an intimate image of a person (the depicted person) is provided on:

- (a) a social media service; or
- (b) a relevant electronic service; or
- (c) a designated internet service;

the provision of the intimate image on the service is an exempt provision of the intimate image if:

(d) the provision of the intimate image on the service is necessary for, or of assistance in:

- (i) enforcing a law of the Commonwealth, a State or a Territory; or
- (ii) monitoring compliance with, or investigating a contravention of, a law of the Commonwealth, a State or a Territory; or

(e) the provision of the intimate image on the service is for the purposes of proceedings in a court or tribunal; or

(f) the provision of the intimate image on the service is for a genuine medical or scientific purpose; or

(g) an ordinary reasonable person would consider the provision of the intimate image on the service acceptable, having regard to the following matters:

(i) the nature and content of the intimate image;

(ii) the circumstances in which the intimate image was provided on the service;

(iii) the age, intellectual capacity, vulnerability or other relevant circumstances of the depicted person;

(iv) the degree to which the provision of the intimate image on the service affects the privacy of the depicted person;

(v) if the intimate image was posted on the service by an end-user of the service—the relationship between the end-user and the depicted person;

(vi) whether the depicted person has died, and if so, how much time has elapsed since the depicted person's death;

(vii) any other relevant matters; or

(h) the following conditions are satisfied:

(i) the intimate image was posted on the service by an end-user of the service;

(ii) the end-user is a protected person (within the meaning of section 223);

(iii) the post was in connection with the exercise of a power, or the performance of a function, conferred on the Commissioner by or under this Act; or

(i) a condition determined under subsection (2) is satisfied.

(2) The Minister may, by legislative instrument, determine one or more conditions for the purposes of paragraph (1)(i).

(Note: Section 223 sets out a list of "protected persons", including the Commissioner, member or staff of the Australian Communications and Media Authority, etc.)

2) Child Abuse Material

34. The Consultation Paper proposes to include online child abuse material “depicting the victim”. This is defined as:

- (a) a person who is or appears / implied to be below 16 years of age (i) as a victim of torture, cruelty or physical / sexual abuse, (ii) engaged in or apparently engaging in a sexual pose or activity, or (iii) in the presence of another person who is engaged in or apparently engaging in a sexual pose or activity; or
- (b) the intimate image of a person who is or appears / implied to be below 16 years of age.³⁹

35. We make the following comments:

- (a) ***Include child abuse material “appearing to depict the victim”.*** By including child abuse material “appearing to depict” victims, this would extend the scope to cover deepfakes, AI-generated and doctored images.⁴⁰ In this regard, we note that the phrase “appearing to depict the victim” appears in the definition of intimate image abuse in the Consultation Paper.
- (b) ***Widen the scope beyond victims.*** Children and youths may not have the maturity or be in a suitable position to take action against the harm in question. We would recommend expanding the scope of whom may make complaints on their behalf, to include schools or other institutions having care of children (see paragraph 49).

3) Online Statements Instigating Disproportionate Harm (“OSIDH”)

36. The Consultation Paper proposes to include “online statement instigating disproportionate harm” (“OSIDH”) among the online harms. OSIDH is defined as online material that:

- is published or intended to be published;
- instigates or purports to instigate the public in Singapore or a section of the public in Singapore to act, or omit to act in a way to cause harm to the victim; and
- is unjustifiable.⁴¹

37. This addresses cancel campaigns, and we are supportive of the move in principle.



³⁹ Annex C, Consultation Paper.

⁴⁰ In the incident involving deepfake nude photos of Singapore Sports School students, it is unclear from reports whether the victims were under the age of 16. However, in the event that there were any under-16 victims, such a definition of online harms could afford the necessary protection. (“Police investigating deepfake nude photos of Singapore Sports School students” *The Straits Times* (13 November 2024): <https://www.straitstimes.com/singapore/police-investigating-deepfake-nude-photos-of-singapore-sports-school-students>)

⁴¹ Annex C, Consultation Paper.

38. As mentioned above, the precise harm done by cancel culture is the *disproportionate penalty of social exclusion, including any related material loss (for real or alleged wrongdoing)*.

39. We raise the following by way of comment:

(a) **Harm.** As the central harm of cancel culture is the *disproportionate penalty of social exclusion*, we would recommend creating a new category of general damages for such a loss that can be awarded in the context of the OSIDH statutory tort. This is on top of established categories of damages such as physical or psychiatric harm, reputational damage, or economic losses.

(b) **Causation.** The consideration of causation is particularly relevant to OSIDH as a statutory tort. Cancel campaigns have sometimes been likened to a “mob”,⁴² making it often unclear whether people are causing harm to the victim *as a result of any one person’s instigation*, or simply as part of groupthink or the frenzy of the moment.⁴³ Nevertheless, we recommend that proof of causation of harm should *remain* as a requirement in the context of the statutory tort, in order to successfully claim damages resulting from that harm. Suitable provisions can be made in law to allow for courts to apportion blame or award a proportionate amount of damages relative to the defendant’s contribution to the claimant’s loss.

(c) **Justifiability.** The concept of “justifiability” could be better analysed as a separate concept from the *proportionality* of those acts. From a legal standpoint, there should be distinctions drawn between illegal acts (which are categorically unjustifiable) on one hand, and legal or legally protected acts on the other (which are justifiable to the extent they are proportionate). We cite the following examples by way of illustration:

Legality of Acts	Justifiability	Example(s)
Illegal	Unjustifiable	<ul style="list-style-type: none"> - A person instigates the public to send death threats to a celebrity for expressing views they disagree with. - A group of people spread false and defamatory rumours about



⁴² As Rabbi Jonathan Sacks wrote, “once the mob has been let loose, it becomes difficult to distinguish between genuine cases of wrongdoing, and other accusations motivated by malice, or a desire for revenge, or some other less than fully moral cause”. (Jonathan Sacks, *Morality: Restoring the Common Good in Divided Times* (London: Hodder & Stoughton, 2020 at 214))

⁴³ For example, in one instance of a cancel campaign against Wendy Cheng (also known as “XiaXue”), a woman who led the campaign to “demonetise XiaXue” retracted her petition and said “I do not stand by some of the things that I said in the heat of the moment”. (“I regret this’: Woman who led ‘demonetise Xiaxue’ campaign retracts all statements made” *AsiaOne* (5 November 2020), online: <https://www.asiaone.com/digital/i-regret-woman-who-led-demonetise-xiaxue-campaign-retracts-all-statements-made>.)

Legality of Acts	Justifiability	Example(s)
		an individual to get him fired from his job.
Legal	Justifiable to the extent they are proportionate	<ul style="list-style-type: none"> - An organisation calls for a speaker to be disinvited from a forum on account that he sexually harassed his ex-colleagues. - A group of people express their disapproval of a collaboration between two companies.
Legally protected (e.g. freedom of association, freedom of religion)	Justifiable to the extent they are proportionate	<ul style="list-style-type: none"> - An aggrieved member of a members-only club calls for the chairperson of the club to be voted out at the next election. - A faction within a religious organisation calls for a leader to be expelled for promoting unorthodox teachings and practices.

(d) **Proportionality.** We are of the view that proportionality should be a key and separate element of OSIDH. This should be analysed by balancing the degree and extent of harm caused to the victim of OSIDH, as opposed to the degree and extent of harm caused by the victim’s real or alleged wrongdoing (if any).

(e) **Government action.** The scope of OSIDH should not include government action. On one hand, petitions to the Government are a legitimate part of the democratic process, so long as they do not fall foul of any other laws. On the other hand, where the Government exercises any of its powers to prohibit, arrest, prosecute anyone and/or otherwise enforce any laws or policies, these should be addressed under the normal legal frameworks (including the Constitution).

40. By way of example, we cite the 2014 incident involving British expatriate Anton Casey, to analyse how OSIDH may apply to a case like his (should it arise again):

Case Study: Anton Casey (2014)

In 2014, British expatriate Anton Casey stirred controversy when he posted a picture of his son on the Mass Rapid Transit (“MRT”) train with the caption “Daddy where is your car and who are all these poor people?” He later posted a picture of his son in a silver Porsche with the caption: “Normal service can resume, once I have washed the stench of public transport off me.”

Condemnation was swift. Casey was doxed by a Facebook page which revealed where Casey worked, who his bosses were, gave e-mail addresses for the public to spam their mailboxes and

even revealed his home address.⁴⁴ Casey's apology through a public relations firm was rejected as insincere, and his lawyer hired to get social media sites to take down certain posts and pictures was subject to name-calling and received nasty emails.⁴⁵

Then-Member of Parliament ("MP") Zainal Sapari wrote in a Facebook post, alluding to the incident with Amy Cheong who was fired from NTUC for racially offensive remarks: "NTUC was decisive in sending a clear message when our staff crossed the line. Will Crossinvest Asia do the same?"⁴⁶

A petition was posted on Change.org with a xenophobic bent, telling Casey to "go back to where you came from" and his Singaporean wife Bernice Wong to "go back to where you should have been born".⁴⁷

Casey was fired from his job at Crossinvest Asia, reportedly received death threats,⁴⁸ and left Singapore amidst threats made to his family.⁴⁹ He hoped to return, and offered to "volunteer [his] time and resources to community projects in order to make amends for [his] mistakes".⁵⁰

There were those who argued that Casey committed the offence of sedition,⁵¹ even as others criticised the disproportionality of the attacks and called for action to be taken against those who threatened Casey's family.⁵² Then-Minister for Foreign Affairs and Law K. Shanmugam



⁴⁴ William Wan, "Anton Casey case: Where has all our empathy gone?" *The Straits Times* (24 January 2014), online: <https://www.straitstimes.com/singapore/anton-casey-case-where-has-all-our-empathy-gone>.

⁴⁵ "Anton Casey loses job, friends" *The New Paper* (28 January 2014), online: <https://www.asiaone.com/singapore/anton-casey-loses-job-friends>.

⁴⁶ "Anton Casey and family leave for Perth amid threats" *The Straits Times* (25 January 2014), online: <https://www.straitstimes.com/singapore/anton-casey-and-family-leave-for-perth-amid-threats>.

⁴⁷ "Go back to where you came from. For Bernice, go back to where you should have been born and where you will like (tall, big, masculine 'mature' men)." *Change.org*, online: <https://www.change.org/p/repatriate-anton-casey-family-back-to-uk-go-back-to-where-you-came-from-for-bernice-go-back-to-where-you-should-have-been-born-and-where-you-will-like-tall-big-masculine-mature-men>.

⁴⁸ "British expat banker Anton Casey causes uproar in Singapore after mocking 'poor people' calling a taxi driver a 'retard'" *The Independent* (22 January 2014), online: <https://www.independent.co.uk/news/business/news/british-expat-anton-casey-causes-uproar-singapore-after-mocking-poor-people-calling-taxi-driver-retard-907795.html>; "British banker receives death threats for anti-Singapore diatribe" *The Guardian* (23 January 2014), online: <https://www.theguardian.com/world/2014/jan/23/banker-singapore-insults>.

⁴⁹ "Anton Casey and employer Crossinvest Asia have 'parted ways'" *The Straits Times* (25 January 2014), online: <https://www.straitstimes.com/singapore/anton-casey-and-employer-crossinvest-asia-have-parted-ways>; "Anton Casey and family leave for Perth amid threats" *The Straits Times* (25 January 2014), online: <https://www.straitstimes.com/singapore/anton-casey-and-family-leave-for-perth-amid-threats>.

⁵⁰ "Anton Casey and family leave for Perth amid threats" *The Straits Times* (25 January 2014), online: <https://www.straitstimes.com/singapore/anton-casey-and-family-leave-for-perth-amid-threats>.

⁵¹ See, for example, Steve Wang, "Didn't Anton Casey commit sedition?" *TODAY* (28 January 2014), online: <https://www.todayonline.com/voices/didnt-anton-casey-commit-sedition>.

⁵² Darius Lee, who is now the Executive Director of Cultivate SG, wrote in his personal capacity then to criticise the threats and disproportionality of the attacks against Casey and his family. (Darius Lee, "Act against threats to Casey family" *TODAY* (27 January 2014); Darius Lee, "Anton Casey did not commit sedition" *TODAY* (31 January 2014)).

condemned Casey's remarks as "Deeply offensive, wrong, and unacceptable", but also "[hoped] that Singaporeans will not attack or flame his family because of his actions".⁵³

41. Analysing the incident through the lens of OSIDH, the petition posted on Change.org could potentially fall foul of the law, due to its xenophobic instigation and disproportionate nature to Casey's wrongdoing. On the other hand, the statement of Minister K. Shanmugam neither instigates anyone to cause harm, nor are his statements unjustifiable. Although the statement of then-MP Zainal Sapari arguably instigates Crossinvest Asia to cause "harm" (i.e. firing Casey), it is arguably justifiable in a legal sense.⁵⁴

4) Hate Speech

42. The Consultation Paper proposes to include "hate speech" among the online harms. The term "hate speech" is defined as "Online communication to the public that incites or is likely to incite feelings of enmity, hatred, ill-will or hostility against, or contempt for or ridicule of a group or a member of the target group in Singapore by reason of that person's membership in the target group." The "target group" will be defined by "a closed list of characteristics". "Hate speech" includes violence-inciting content, and a "target group" in this latter context can be defined by "any characteristic".⁵⁵
43. We note that the definition of "hate speech" in the Consultation Paper appears similar to the existing offences under the Maintenance of Religious Harmony Act ("MRHA") as well as the proposed offences under the upcoming Maintenance of Racial Harmony Bill.⁵⁶
44. The MRHA adopts different legal thresholds in relation to urging violence on one hand, and of insulting or ridiculing a religion on the other, with a wider definition of a "target group" in the former:
- (a) Section 17E – Offence of urging violence: The "target group" is defined according to "ethnicity, descent, nationality, language, political opinion or any other characteristic (whether or not of a similar kind)".⁵⁷



⁵³ K Shanmugam SC, Facebook post at 10.13am on 23 January 2014, online: <https://www.facebook.com/k.shanmugam.page/posts/646349898744877>.

⁵⁴ While we are concerned about the phenomenon of cancel culture as a whole, we are mindful of the limits to which laws against online harms can address situations where there are calls for people to be fired from their jobs.

⁵⁵ Annex C, Consultation Paper.

⁵⁶ Sections 17E to 17G, Maintenance of Religious Harmony Act; Ministry of Home Affairs, "Public Consultation on the Maintenance of Racial Harmony Bill" (16 April 2024): <https://www.mha.gov.sg/mediaroom/press-releases/public-consultation-on-the-maintenance-of-racial-harmony-bill/>.

⁵⁷ Sections 17E(1)(d) and (2)(c), Maintenance of Religious Harmony Act.

- (b) Section 17F – Offence of inciting hatred, ill-will, etc.: The “target group” is one which is “distinguished by religion or religious belief or activity”.⁵⁸

45. We make the following suggestions:

- (a) **Definition of the “target group”.** We suggest that the definition of a “target group” should adopt the exact same structure as the MRHA as well as, potentially, the upcoming Maintenance of Racial Harmony Bill. In other words, a wider definition of a “target group” should apply in relation to incitement of violence so that it covers any group of persons based on any identifiable characteristic,⁵⁹ whereas a tighter definition of a “target group” – limited to the established categories of race and religion⁶⁰ – should apply in relation to inciting hatred, ill-will, etc. (i.e. offensive content).
- (b) **Objective standard, as opposed to subjective standard.** Terms such as “feelings of enmity, hatred, ill-will or hostility” and “contempt for or ridicule of” are potentially subjective in scope and interpretation. We recommend that the Agency should adopt an objective standard in assessing these harms, taking into account all factors including an objective assessment of the context of the online content in question. This would better guard against the risk that individuals may subjectively manufacture and or trump up outrage to censor perspectives they disagree with.
- (c) **Relevance (or otherwise) of individual moral culpability.** The role of the Agency is not that of a court in a criminal prosecution. Thus, individual moral culpability or *mens rea* (e.g. “intention”, “knowledge”, etc.) would be less relevant. We would suggest that, as long as the acts constituting the online harms meet the requisite threshold for action, the Agency should be empowered to take the necessary measures such as access blocking orders, regardless of individual *mens rea*.
- (d) **Provide illustrations and examples.** To give further clarity in relation to the scope of permissible and prohibited forms of online speech or conduct, we recommend the insertion of illustrations and examples inserted into the law that set out various scenarios where the law is violated or not. Such illustrations and examples are not only useful for the Agency (and courts) when interpreting the scope of the provision,⁶¹ but are also instructive to the public.



⁵⁸ Sections 17F(1)(b) and (3)(b), Maintenance of Religious Harmony Act. See also the definitions of a “target person” in the section.

⁵⁹

⁶⁰ In a 2019 speech, Minister for Law and Home Affairs K. Shanmugam described “racial and religious harmony” as part of the “fundamental assurance one gets in Singapore”. (Ministry of Home Affairs, “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” (1 April 2019): <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/>)

⁶¹ See, for example, the case of *Lee Chez Kee v Public Prosecutor* [2008] 3 SLR(R) 447 where the Court of Appeal considered the illustration under the Evidence Act in its interpretation of the admissibility of certain evidence.

5) Recommendation: Other Harms to be Included

46. We recommend that the following be included in the list of online harms addressed by the Agency:

(a) ***Sexualised communications to children.*** Sections 376EA to 376EE of the Penal Code presently prohibit sexual grooming, sexual communication, and sexual activity or images in the presence of minors.⁶²

(b) ***Advocating or instructing on suicide or self-harm.*** Section 45D of the Broadcasting Act includes “content that advocates or instructs on suicide or self-harm” within the definition of “egregious content” that can be blocked from access.⁶³ Section 306 of the Penal Code criminalises the abetment (which is defined to include instigation and aiding) of suicide.⁶⁴

6) Other Considerations: Vulnerable Persons

47. There are persons who may be more vulnerable than others to online harms. These include children and young persons, the elderly, those who lack sufficient mental capacity and people with mental health conditions. Due to lack of agency, unfamiliarity with technology, various impairments or other reasons, it may be difficult for them to access the reliefs provided by the law despite their heightened risk.

48. We note that, under the Consultation Paper, the scope of whom may subject a complaint to the proposed Agency is as follows:

(a) ***Victims.*** Victims allowed to submit a complaint should be Singapore citizens or Permanent Residents, or foreigners with a student’s pass, work permit or work visa (including dependents who have the applicable pass). Corporate entities based in Singapore will also be allowed to submit complaints.⁶⁵

(b) ***Victims’ representatives.*** These include the parents or guardians of victims under 18 years of age, and individuals who have been authorised by the victim in writing to submit a complaint to the Agency on their behalf.⁶⁶



⁶² Sections 376EA to 376EE, Penal Code. This could potentially address situations such as the one which took place during the COVID-19 pandemic when hackers allegedly hijacked students’ Zoom home-based learning to show pornographic content, and asked the girls to flash their chests. (“MOE suspends use of Zoom for home-based learning after hackers hijack classes” *TODAY* (10 April 2020): <https://www.todayonline.com/singapore/moe-suspends-use-zoom-home-based-learning-after-hackers-hijack-classes>)

⁶³ Section 45D, Broadcasting Act.

⁶⁴ Sections 107 and 306, Penal Code.

⁶⁵ Consultation Paper at 4.

⁶⁶ Consultation Paper at 4.

49. In order to better protect vulnerable persons, we would invite the Government to consider the following measures:

- (a) ***Expand the scope of third parties allowed to make complaints to the Agency on behalf of victims.***⁶⁷ We recommend expanding the scope of whom may make complaints to include representatives from institutions having care of such vulnerable persons (e.g. schools, mental health institutions, family service centres, etc.), who are permitted to report online harms to the Agency for necessary action.⁶⁸
- (b) ***Allow the Agency to consider complaints of other third parties on a case-by-case basis.*** In addition to (a) above, we would recommend allowing the Agency to consider the complaints of other (non-victim) third parties who make complaints about harmful online content, albeit on a case-by-case basis. Such measures may be particularly relevant in the context of child abuse materials (including CSAM), where there is strong public interest in favour of stamping out the circulation of such content.
- (c) ***Empower the Agency to take into account special vulnerabilities of victims.*** In assessing whether the online harm is sufficiently severe to warrant action by the Agency, the Agency should be allowed to take into account any special vulnerabilities of victims. For example, a person with suicidal ideation may merit special protections against harassment or abetment of suicide (e.g. the incident involving Charlotte Dawson).⁶⁹


⁶⁷ For example, in the incident at Singapore Sports School involving deepfake nude photos of students, it was the school that lodged a police report. (“Police investigating deepfake nude photos of Singapore Sports School students” *The Straits Times* (13 November 2024): <https://www.straitstimes.com/singapore/police-investigating-deepfake-nude-photos-of-singapore-sports-school-students>)

⁶⁸ Another possible alternative to this may be to make it a standard practice for institutions having care of vulnerable persons to obtain a signed authorisation from such persons or their legal representatives, so that the institutions may lodge complaints to the Agency on such persons’ behalf. This practice could also be adopted by social service agencies or counselling organisations, so that they may lodge complaints to the Agency on behalf of their clients.

⁶⁹ “Charlotte Dawson’s death puts cyberbullying back in spotlight” *Australia Broadcasting Corporation* (23 February 2014): <https://www.abc.net.au/news/2014-02-23/charlotte-dawson-death-puts-focus-on-cyber-bullying/5277904>. Conversely, empowering the Agency to take into account special vulnerabilities of victims will also prevent over-enforcement in cases where such vulnerabilities do not exist.

B. Role of the Proposed Agency on Online Harms

50. We support the proposal to establish the Agency, which is dedicated to supporting victims of online harms and enhance online safety.
51. Among the positive aspects of the Agency, which we appreciate, are:
- (a) ***Speedier responses.*** The dedicated Agency allows for speedier responses to various forms of online harms, including the “Access blocking order” and “App removal order”, which must be complied with within a specified time. This allows for the harm to be stopped quickly, as compared to the traditional legal process.
 - (b) ***Psychological and practical value to victims.*** Having an Agency to address the online harms, instead of leaving it to individuals to pursue the matter in the courts, allows for some degree of psychological and practical relief to the victims’ distress.
52. We set out some comments on two aspects of the proposal, concerning the publication of the Agency’s orders, and avenues for appeal.
- 1) Publication of the Agency’s Orders
53. The Consultation Paper does not mention whether or not the Agency will publish some or all of the orders it makes, and in what manner of detail (if any).⁷⁰
54. We are of the view that the Agency’s duty to publish or refuse publication of orders should be explicitly stated in the law. Among the considerations the Agency should take into account include the nature of the online harm, vulnerability of the victim, interests of privacy, reputational damage, and overriding public interests. The Agency should also be under a duty to redact such details of persons as may be necessary to protect victims or the public, in the event of publication.
55. This is because, in the context of different online harms, there may be competing considerations:
- (a) ***Vulnerable victims.*** Vulnerable victims such as children and young persons, the elderly, those who lack sufficient mental capacity and people with mental health conditions may merit special privacy protections. In the context of certain harms such as intimate image abuse, victims (both women and men) are also a vulnerable group.
 - (b) ***Publication may be harmful in cases involving invasions of privacy.*** In categories of online harms which involve invasions of privacy (such as harassment, intimate image abuse and misuse of personal information), further



⁷⁰ By comparison, the POFMA Office, responsible for the administration of the Protection from Online Falsehoods and Manipulation Act (“POFMA”), publishes its directions and declarations online in some detail: <https://www.pofmaoffice.gov.sg/media-centre/>.

publicity may aggravate harm to the victim. Malicious actors may even desire such publicity in order to cause or aggravate the harm to victims.

- (c) ***Publication may be beneficial in cases involving falsehoods and reputational damage.*** There are categories of online harms that may in some cases benefit from publication of orders made by the Agency (such as false statements and statements affecting reputation). In OSIDH, publication may be beneficial in some cases where the harmful online material has already been made public or in wide circulation, in order to counter the harm.
- (d) ***Overriding public interest may apply in certain cases.*** Public interest may be an overriding consideration, such as in harms affecting racial and religious harmony. Limited publication of orders may be warranted when addressing racially and religiously incendiary remarks which fall within the scope of “hate speech” under the proposed measures. On one hand, the inflammatory and possibly widely-circulated nature of the remarks may require a public denunciation to counter the harm. On the other hand, it may be against the public interest for the Agency repeat the remarks in totality, in order to avoid giving the remarks further publicity.⁷¹

2) Avenues for Appeal

- 56. Judicial oversight of administrative bodies is an important aspect of the rule of law in Singapore.⁷² This includes the proposed Agency.
- 57. Currently, a person who is dissatisfied with directions or declarations made under the Protection from Online Falsehoods and Manipulation Act (“POFMA”) may proceed to appeal against them as follows:
 - (1) Firstly, by applying to the relevant Minister to vary or cancel the POFMA direction or declaration,⁷³ using a standard form which is provided online. The Minister must issue a notice of his or her decision within 2 working days; if the Minister does not do so, the application is treated as refused.⁷⁴


⁷¹ For an example of such limited publication concerning sensitive remarks, please refer to the police press statement concerning the remarks made by Subhas Nair on race and religion. (Singapore Police Force, “Man To Be Charged For Attempts To Promote Feelings Of Ill-Will Between Different Groups On Grounds Of Religion And Race” (28 October 2021: https://www.police.gov.sg/media-room/news/20211028_man_to_be_chrg_fr_attmpt_to_prmte_feelngs_of_illwill_btwn_diff_grp_on_grnd_of_relgion_n_rce)

⁷² The Court of Appeal has opined that one of the “core ideas” of rule of law is the notion that “the power of the State is vested in the various arms of government and that such power is subject to legal limits... Judges are entrusted with the task of ensuring that any exercise of state power is done within legal limits.” (*Tan Seet Eng v Attorney-General and another matter* [2016] 1 SLR 779 at para. 1)

⁷³ Sections 19, 31, 32 and 46, Protection from Online Falsehoods and Manipulation Act.

⁷⁴ Regulation 14, Protection from Online Falsehoods and Manipulation Regulations.

(2) Secondly, to appeal to the General Division of the High Court within 14 days after the application to the Minister to vary or cancel the POFMA direction or declaration is refused.⁷⁵

(3) If the General Division of the High Court rejects the application, further appeal to a higher court may be possible in certain cases.⁷⁶

58. We are of the view that a similar process should be available for appeal against decisions of the Agency.

C. Improving Accountability in the Online Space

59. Cultivate supports the proposal to increase online accountability through improved user information disclosure.⁷⁷ In this regard, it is important to balance the interests of privacy on one hand, and prevent the abuse of anonymity to perpetrate online harms on the other.

60. Internet anonymity has its positive and negative sides. One positive side of internet anonymity is, for example, its value in encouraging people with mental health struggles to seek help by minimising fear or stigma for the people affected.⁷⁸

61. However, the “dark side” of internet anonymity poses a well-known problem to existing legal frameworks, whether civil or criminal. Civil wrongdoing, such as defamation or harassment, can be perpetrated anonymously. Since identifying a defendant is critical in a civil suit, an inability to identify the potential party to be sued may effectively prevent a victim from pursuing a civil claim.⁷⁹

62. In order to pierce the anonymity of the internet and ascertain the proper party to be sued in a civil claim, a party may have to sue the service provider to compel it to disclose the identity of the person behind the anonymous online posts.⁸⁰ This necessitates two separate (and potentially long-drawn) actions in court:

⁷⁵ Rule 5, Supreme Court of Judicature (Protection from Online Falsehoods and Manipulation) Rules.

⁷⁶ Rule 16, Supreme Court of Judicature (Protection from Online Falsehoods and Manipulation) Rules.

⁷⁷ Consultation Paper, paras. 31 to 32.

⁷⁸ See, for example, “Groups offer online counselling to young people who need help but wish to remain anonymous” *The Straits Times* (12 December 2022): <https://www.straitstimes.com/singapore/groups-offer-online-counselling-to-young-people-who-need-help-but-wish-to-remain-anonymous>.

⁷⁹ The case of *Qingdao Bohai Construction Group Co, Ltd and others v Goh Teck Beng and another* [2016] 4 SLR 977 clearly illustrates this point, where the plaintiff was hindered in its claim because of the anonymity of the publisher(s) of certain articles which allegedly contained untrue, scurrilous and defamatory statements.

⁸⁰ In *Kabbabe v Google LLC* [2020] FCA 126, a dental surgeon Dr Kabbabe sued Google in order to identify an unknown person who posted an allegedly defamatory review in relation to Dr Kabbabe’s dental practice on Google, using a pseudonym “CBsm 23”. He was successful in obtaining an order for preliminary discovery. (See also *Intas Pharmaceuticals Ltd v DealStreetAsia Pte Ltd* [2017] 4 SLR 684,

- (1) First, the party must commence pre-action discovery against a network service provider, social media company or the platform where the post has been made, in order to obtain the identity of the user (or documents that may lead to the identification of the user);
 - (2) Second, assuming the first step is successful and the user can be identified, the party can then proceed with such civil claims against that user who made the anonymous post.
63. Not only are such procedures costly, success is also not guaranteed for various reasons, such as opposition from the tech company (which is generally better-resourced than many litigants), failure to reach the requisite legal thresholds for pre-action discovery,⁸¹ or the service provider may itself be unable to identify the internet user.⁸² Precious time is also lost in the process, all while the victim continues to suffer the consequences of the online harm. Internet anonymity (when abused to perpetrate online harms) thus poses a significant barrier to justice to a person who is seeking recourse for civil wrongdoing.⁸³

1) Disclosure of User Information

64. We note that the Consultation Paper proposes to allow a victim who has filed a complaint with the Agency to “apply to the Agency for the disclosure of a perpetrator’s user information for certain specified purposes (for example, to bring a claim under the statutory torts or to safeguard themselves from the perpetrator)”. It adds that the Agency may disclose the information to the victim, subject to the victim satisfying prescribed requirements.⁸⁴
65. We are of the view that, in principle, the scope of disclosure of user information should be the same or of a similar position as currently available under civil law in relation to pre-action discovery, including the *Riddick* principle (or implied undertaking) that documents or information disclosed should not be used for extraneous purposes.⁸⁵ Breach of the *Riddick* undertaking is contempt of court, and punishable accordingly.⁸⁶



where pre-action discovery was sought against a financial news website in order that the plaintiff could plead malice and commence claims against certain unnamed sources for malicious falsehood.)

⁸¹ The process of pre-action discovery is intended “to accommodate the situation where a potential plaintiff does not have sufficient facts to commence proceedings”, and is guided by the principle of “necessity”. (*Ching Mun Fong v Standard Chartered Bank* [2012] 4 SLR 185 at para. 23)

⁸² In the context of copyright claims, see Cheah Yew Kuin and Faith Lim, “Revisiting Pre-action Discovery: Applications against Network Service Providers for Subscriber Details” *Law Gazette* (December 2016), online: <https://v1.lawgazette.com.sg/2016-12/1724.htm>.

⁸³ For an illustration of the legal difficulties involved, see the case of *Qingdao Bohai Construction Group Co, Ltd and others v Goh Teck Beng and another* [2016] 4 SLR 977.

⁸⁴ Consultation Paper, para. 32.

⁸⁵ See, for example, *ED&F Man Capital Markets Ltd v Straits (Singapore) Pte Ltd* [2020] 2 SLR 695 at para. 66.

⁸⁶ *Pertamina Energy Trading Ltd v Karaha Bodas Co. LLC* [2007] 2 SLR(R) 518 at para. 38.

66. To improve accountability and render justice to victims of online harms, we recommend the introduction of two forms of disclosure:

- (a) **Limited disclosure.** Limited disclosure is intended to allow victims to identify the perpetrators of online harms. The threshold to obtain such information should be similar to the legal threshold to establish civil wrongdoing on a *prima facie* basis. Such disclosure should be subject to the condition that the information obtained cannot be used for any extraneous purpose.⁸⁷ Violation of this condition should be subject to criminal penalties, the severity of which should be pegged at a similar level to contempt of court. Our proposal here is similar to the current proposal under the Consultation Paper.
- (b) **Full disclosure (remedy which can be granted by courts).** In addition, we would recommend adding the option of full disclosure of user information as a possible remedy that can be granted by a court. This is intended to fully expose the anonymous, pseudonymous or fake account which has been used to perpetrate online harms, in order to promote accountability and redress the wrong done. Here are some examples of when a full disclosure order may be warranted:

Category of Harm	Example(s)
Impersonation	<p>A perpetrator (P) impersonates a victim (V) to post numerous offensive and defamatory comments against various people. As a result, V is subjected to harassment and threats of lawsuits.</p> <p>V wishes to publicly expose P as the person behind the various offensive and defamatory comments, in order that the harassment and threats of lawsuits against V will cease.</p>
<p>Online statements instigating disproportionate harm (“OSIDH”)</p> <p>False statements</p> <p>Statement affecting reputation</p>	<p>A perpetrator (P) creates multiple “news sites” spreading false and defamatory rumours about a victim (V) while calling for a boycott of V’s business, which are circulated widely on the internet.</p> <p>V wishes to publicly expose P as the person behind all these “news sites”, to show that the same person P is pursuing a personal vendetta against V.</p>



⁸⁷ This is similar to the *Riddick* principle in law. “The *Riddick* principle states that a party who discloses a document in discovery in an action under compulsion is entitled to the protection of the court against any use of the document otherwise than in that action...” (*ED&F Man Capital Markets Ltd v Straits (Singapore) Pte Ltd* [2020] 2 SLR 695 at para. 66)

2) Platforms' Collection of User Information

67. We would presume that, in order for the operation of this area of law to be effective, online platforms may be required to collect a certain amount of user information when users sign up on those platforms. This is in order that users can be sufficiently identified for legal purposes should they abuse the platforms to perpetrate online harms.
68. We are supportive of such a position in principle, subject to sufficient protection for personal data.
69. We raise the following by way of comment:
- (a) ***Scope of extent of user information to be collected.*** As victims of online harms are most likely to need sufficient information in order to pursue civil claims against the perpetrators (including the new statutory torts), the types and amount of user information collected should be sufficient for such purposes. Currently, the information required under the Rules of Court and Practice Directions for a civil suit are the respective parties' **(i) names** and **(ii) identification numbers**.⁸⁸ For the purposes of service of court documents, **(iii) location** and **(iv) relevant contact details** are necessary, including email or instant messaging handles (for substituted service).⁸⁹
 - (b) ***Relaxation of rule requiring identification numbers in court documents.*** The current Rules of Court and Practice Directions require parties' identification numbers to be inserted in court documents as part of a civil suit.⁹⁰ Given that many online platforms may not collect identification numbers, we would recommend relaxing this rule in relation to the commencement of civil suits, at least in relation to online harms.



⁸⁸ See Forms 8 and 15 under Appendix A of the Supreme Court Practice Directions, in relation to originating claims and originating applications respectively. In this regard, we note – but express no opinion on – the recent controversy regarding the disclosure of NRIC numbers on the Accounting and Corporate Regulatory Authority (“ACRA”) portal. (“Acra disables search function for NRIC numbers on portal for now” *The Straits Times* (14 December 2024): <https://www.straitstimes.com/singapore/acra-disables-search-function-for-nric-numbers-on-portal-for-now>)

⁸⁹ Where the party has a Singpass account, substituted service through the Singpass app may also be a possibility. (SG Courts, “Media Release: New electronic option to effect substituted service of court documents for civil proceedings” (17 August 2022): <https://www.judiciary.gov.sg/news-and-resources/news/news-details/media-release-new-electronic-option-to-effect-substituted-service-of-court-documents-for-civil-proceedings>)

⁹⁰ See Forms 8 and 15 under Appendix A of the Supreme Court Practice Directions, in relation to originating claims and originating applications respectively.

- (c) **Third-party verification of accuracy.** We would invite the Government to explore the possibility of third-party verification of the accuracy of user information, rather than to mandate that platforms directly collect and store sensitive personal information such as images of NRIC cards.⁹¹ A possible candidate to provide such third-party verification is Singpass, which may additionally use a 2-step verification process as an additional safeguard against impersonation. This third-party verification process can potentially also be used for age-verification purposes.

Scenario	Description	Pros and Cons
Platform directly collects and stores personal information	<p>A new user (U) creates an account with an online platform (P), which requires U to provide U's (i) name, (ii) identification number, (iii) location and (iv) contact details including email address and phone number.</p> <p>U provides the information.</p> <p>To verify the information, P requires U to take a photograph of U's NRIC for verification purposes.</p>	<ul style="list-style-type: none"> - The platform (P) stores a copy of the NRIC of the user (U), which raises privacy concerns. - There is a risk of inaccuracy or impersonation, since the user (U) can assume a false identity and upload a fake image of an NRIC, or impersonate someone else by using that person's NRIC.



⁹¹ The Government distinguished between the significance of the NRIC card and NRIC number as means of authentication. While the Government has likened an NRIC number to a person's full name, an NRIC card "contains other information such as my photo and fingerprint, that allows others to check that my NRIC card matches me, the person holding the card" (Ministry of Digital Development and Information, "Closing Remarks by Minister Josephine Teo at Press Conference on Responsible Use of NRIC Numbers" (19 December 2024): <https://www.mddi.gov.sg/closing-remarks-by-minister-josephine-teo-at-press-conference-on-responsible-use-of-nric-numbers/>)

Scenario	Description	Pros and Cons
<p>Third-party verification of accuracy of user information (Using the example of Singpass as a third party verifier)</p>	<p>A new user (U) creates an account with an online platform (P), which requires U to provide U's (i) name, (ii) identification number, (iii) location and (iv) contact details including email address and phone number.</p> <p>U provides the information.</p> <p>In order to verify the information, P does not collect a photograph of U's NRIC for verification purposes.</p> <p>Instead, P checks with Singpass whether or not the information provided is accurate.</p> <p>Singpass replies with either "Yes" or "No" in relation to the accuracy of the personal data provided by U to P. Singpass does not give P any other information.</p>	<ul style="list-style-type: none"> - The platform (P) does not store a copy of the NRIC of the user (U), but only has the (Singpass-verified) user information provided. - Singpass verifies the accuracy of the user information. - To create additional safeguards against impersonation, a 2-step verification process can be created, where the Singpass app can verify with the NRIC holder if he or she is indeed setting up a new account with the platform (P).

- (d) ***Protection of personal data, by strictly limiting their collection, use and disclosure to legal purposes.*** The collection of user information by platforms should be buttressed by robust protection of personal data, under the Personal Data Protection Act. Among other things, the collection, use and disclosure of personal data must be strictly limited to legally required or permitted purposes.⁹²



⁹² Section 18, Personal Data Protection Act.

IV. Conclusion

70. As a society with high levels of internet connectivity in an age of rapid technological advancement, the online space is virtually ubiquitous in our daily lives, reaching into spaces both public and private. With these capabilities, the challenge for society is to harness the strengths of technology while addressing and preventing its harms.
71. The proposed new measures to enhance online safety are a welcome move. Stronger protections can act as deterrents and signal the values that we wish to uphold as a society. Beyond protecting individual victims, it helps to promote the common good by preserving the overall tone of public discourse and allowing for responsible civic participation.
72. At the same time, Cultivate recognises that legislation alone is insufficient. It will need to be complemented by a better set of moral norms about how the internet can be kept safe for all users, and by the efforts of everyone to help cultivate a better culture online.