AWARE's Response to the Public Consultation Paper on Enhancing Online Safety

Our response to questions from the Ministry of Law and the Ministry of Digital Development and Information on their public consultation paper on enhancing online safety are as follows:

Section A: Government Agency dedicated to supporting victims of online harms and enhance online safety

A1: Scope of online harms within the Agency's oversight

We propose setting up a government agency dedicated to supporting victims of online harms, with powers to help them get timely assistance.

4. Do you agree that this proposal would be useful for victims of online harms?

Yes

5. Please let us know the reasons for your answer.

AWARE has long advocated for such an agency to more immediately and quickly address online harms without the delays attendant with legal actions in court. The success of the Australian eSafety Commissioner in quickly addressing online harms provides evidence that this is a good approach to adopt. We believe it will make a big difference in giving victim-survivors the framework and access they need to address online harms quickly.

We propose that the Agency be empowered to act on specified types of online harms (see paragraph 17 of the detailed consultation paper):

i. Online harassment
ii. Intimate image abuse
iii. Child abuse material
iv. Impersonation
v. Misuse of inauthentic material (e.g. deepfakes)
vi. Online statements instigating disproportionate harm
vii. Hate speech
viii. Misuse of personal information
ix. False statements
x. Statements affecting reputation

6. Do these categories cover the range of harmful online content you are concerned about? Are there other types of harmful online content that should be included in these categories? If so, please share examples of harmful content that do not fall in the above categories.

We agree with the majority of items on this list as the types of online harms which should fall under the purview of the Agency. However, we suggested a few additions and clarifications and some deletions to achieve clarity on the ambit of the Act.

i. Intimate image abuse

We recommend that the definition should be expanded to expressly include the following categories of image-based sexual abuse:

i. Sextortion: the use of images or recordings which are intimate or have a sexualised context as leverage to threaten, blackmail or otherwise solicit an outcome desired by the perpetrator.

ii. Non-consensual communication of sexual images: causing a victim to non-consensually receive images or recordings which are intimate or have a sexualised context, whether of the victim or not.

Illustration: An example would be the non-consensual sending of “dick-pics”.

ii. Online statements instigating disproportionate harm (“OSIDH”)

It is not very clear to us what “ill” this limb is trying to address based on the information provided. The description is too broad and can be capable of misuse if the parameters are not clear. We think the following areas should be clarified:

i. What are the types of “harm” to be covered (for example physical, reputation, emotional, psychological)

ii. The factor of “disproportionate harm” should be reflected in the level of seriousness of harm which will warrant action; the threshold of harm should reflect protection against fairly serious harm; “unjustifiable harm” is unclear as it invites a question as to what is justified online material.

We also recommend that the definition be clarified with illustrations to give examples of this harm.

iii. Statement affecting reputation (“SAR”)

This harm goes very far beyond the scope of the tort of defamation because it

covers factually correct statements and also opinions that lower the estimation of a person.

For example, if A published a statement of accurate fact about B that is likely to cause a reasonable person to have a lower estimation of B, A would be committing a harm under this definition even though they would not have defamed B under the law of defamation because it is a true statement of fact.

We recommend that this harm should only be restricted to false statements and should not extend to factually correct statements and opinions. The internet has been valuable in exposing abuses of power (for example #meToo related cases). Spaces that allow for truth and for issues and matters of genuine concern to be raised should be protected whilst ensuring harms are eradicated. This is an important balance that we feel very strongly about.

As proposed, SAR is extremely wide and susceptible to being used to suppress the truth, erode freedom of speech and police opinions. People should be restrained by the parameters of the law of defamation. We therefore recommend that the SAR should be subject to the principles of law in relation to defamation, including the defences thereto: justification, fair comment and qualified privilege.

This definition is also too broad and should be clarified:

i. The extent of the harm is unclear: What level of “lower estimation” is meant to constitute this harm? The ambit of a lower estimation of a person can range from a minor lowering of estimation to an extremely low view of a person.

ii. It is unclear what constitutes “harm” and its parameters

We therefore recommend that SAR and the harm of “false statements” be combined and that the defences available under defamation apply.

Related to this, we note that there is some cynicism and concerns on the ground that this Act can be used as a political tool by the government against individuals and activists and this provision in particular is of concern. It is thus important for the Agency and the e-Safety Commissioner to be seen to be independent offices. In addition, given that the Protection from Online Falsehoods and Manipulation Act (POFMA) already provides protection for the instances covered thereunder, it may be worth specifying that this Act will not cover false statements that fall within POFMA.

A2: Types of assistance available from the Agency

The Agency will be able to issue directions to communicators (the perpetrators who post harmful content), administrators and/or platforms (see paragraph 20 of the detailed consultation paper). These include requiring the recipient of the direction to take down or disable access to an offending post, or to communicate a reply notice. The Agency will also have investigative powers to facilitate its assessment of complaints.

7. Do you agree that the most important function of the Agency should be to ensure that offending content is removed swiftly and permanently?

Yes.

8.What other powers do you think the Agency should have, to assist victims of online harms?

Here is a case study of the experience of a client of AWARE to illustrate what type of harm can be experienced: The client's video was uploaded on a social media platform without her consent and remained available even five months after she filed a report with the platform. Throughout the process, the lack of clarity about the timeline, specifically when she would hear back from the platform, was extremely traumatising for her. This, combined with the platform's inaction (as the video was not deemed a violation of its policies), resulted in the client expressing suicidal ideation. In addition to the harm she suffered, it is also true that the longer harmful or offensive content stays online, the greater the risk of it being circulated further. Even if these materials are removed by the platform at a later stage, users of these platforms would have had plenty of time to download and further circulate the content either privately or on their social media networks.

Victim-survivors thus live in a perpetual state of fear, not knowing if the content is still being circulated without their knowledge, and whether it will resurface online one day.

Time frame of 24 hours

We therefore recommend that there should be a time frame within which a take down or disable access order issued by the Agency to a communicator, administrator or platform must be complied with. In Australia, under the Online Safety Act 2021, the eSafety Commissioner can order compliance with a removal notice within 24 hours or such longer time frame specified by their eSafety agency.

We recommend a similar approach and the adoption of a 24 hour time frame (or such other time frame specified by the Agency) within which a communicator, administrator or platform must comply with a take down or disable access order.

Temporary take down and disable access orders

In addition, in order to minimise harm to victim-survivors, we also recommend that upon receipt of a complaint by the Agency, the Agency should have the power to order a temporary take down or disable access order whilst the Agency carries out

its investigations, subject a specified limit of time, which can be renewed by the Agency. This should be based on the expected time taken for investigations and the resources expected to be available to the agency. This would ideally be no more than a month to three months, to encourage swift resolution of complaints raised to the Agency.

Measures to promote dispute resolution

Measures that provide for non-contentious resolution between parties would be helpful to promote resolution without having to resort to court actions only.

We recommend the consideration of measures such as an apology law, available in jurisdictions such as the USA, Canada and Hong Kong, which seeks to promote apologies and apologetic discourse as an important form of out-of-court dispute resolution, by making apologetic statements inadmissible for proving liability in civil wrongs. The Apology Law of Hong Kong was the first in Asia and is considered one of the broadest apology laws, which promotes not only statements of remorse, but also statements of facts embedded in apologies. We recommend inclusion of such a measure in the Act.

Mediation is a form of resolution which is common in Singapore. However, given that the type and severity of harm that can be caused by online harms, mediation is not always suitable in all cases as the forced facing of a perpetrator of harm can be re-traumatising for a victim-survivor. We therefore recommend that the Agency provide the opportunity for mediation but mediation should be directed only upon the consent of both parties.

Measures to assist victim-survivors and perpetrators

Given the type of harm that victim-survivors can suffer as a result of online harms, it would be very good to have avenues for victim-survivors to receive some emotional support when they approach the Agency for help. Measures can include counselling and a trauma-informed trained befriender to assist the victim-survivor.

Perpetrators of online harms also may need help to address the underlying causes of the actions. Similar to the orders under the amended Women's Charter (Family Violence and Other Matters) Amendment Act, the Agency could be empowered to order mandatory counseling for perpetrators, subject to satisfaction of the need for such an order.

Training for Agency officials

The Agency will be receiving complaints from victim-survivors who have suffered harm, and many of whom will have suffered trauma, especially where the harm was of a sexual nature. It is very important that the way in which Agency officials handle the victim-survivors is trauma-informed, so as not to aggravate the harm they have suffered and in order to best support them. We there recommend that all Agency

officials who will be dealing with victim-survivors should undergo training on trauma-informed approaches to adopt in their interactions with victim-survivors.

Section B: New statutory torts to clarify (i) the online harms that people can sue for; (ii) the parties whom they can sue; and (iii) the types of relief that they may obtain in court, including damages.

B1: Scope of online harms covered by the statutory torts

We propose that victims can bring a claim in Court for specified types of online harms (see paragraph 26 of the detailed consultation paper):

i. Online harassment (including sexual harassment, doxxing and cyberstalking)
ii. Intimate image abuse
iii. Child abuse material
iv. Impersonation
v. Misuse of inauthentic material
vi. Online statements instigating disproportionate harm
vii. Hate speech (for violence-inciting content only)

9. Do you agree that victims should be allowed to bring an action in Court for these types of online harms?

Yes, but subject to our recommendations under question 6 on the additions, clarifications and deletions that we recommend to the proposed categories of online harm.

In addition, we recommend the expansion of this list of categories of online harm covered by statutory tort to include “Misuse of personal information”. We note the explanation that this category has been excluded on the basis that there could be recourse under existing laws on breach of confidence or breach of contract. However, these laws (breach of confidence and breach of contract) only apply in specific cases, for example where there is contract between the parties or there is an obligation under the law to maintain the confidence of another (usually in the context of a contract). They do not have general application in all the cases when a misuse of personal information could occur online, for example where someone intentionally divulges such personal information out of spite or as a form of mischief or prank. We therefore recommend that the inclusion of this category of harm as a statutory tort.

10. Please explain your reasons.

Please see the answer to question 9.

We are of the view that creating these statutory torts will act as a deterrent to would-be perpetrators of such harm. The anonymity and lack of consequences for causing harm online has enabled perpetrators of online harm to cause abuse without fear of punishment or other personal consequences.

Furthermore, victim-survivors who suffer harm should rightly be entitled to remedies for such harm against the perpetrator of the harm. This is similar to the recourse available in law for other torts such as defamation, assault, battery, and trespass to land and chattels. In all these instances, harm is caused either to the person, property or reputation of the victim and remedies are available under the law for them. Similarly, we believe it is appropriate for these online harms to be designated as statutory torts for which remedies should be available.

B2: Duties owed by communicators, administrators and platforms

Our detailed consultation paper describes the duties imposed on communicators, administrators and platforms, under the statutory torts (see paragraphs 28-30 of the detailed consultation paper).

11. Do you agree that communicators, administrators and platforms should be subject to these duties?

Yes.

However, we note that there is some ambiguity around the following aspects of Annex F:

- Section 6(a): the definition and ambit of the defence of “reasonable conduct”
- Section 6(b): the definition and ambit of the defence of “reasonable excuse”
- Section 6(b): the ambit of the steps required to establish “reasonably assessed”
- Section 6(b): under what circumstances it would be considered “reasonable or appropriate not to take steps or measures”
- Section 6(b): the ambit of the steps required to establish “reasonably practicable”
- Section 11: the steps being considered to establish that they have acted reasonably
- Section 12: the definition and ambit of the defence of “reasonable excuse”
- Section 12(a): the ambit of the steps required to establish “reasonably assessed”
- Section 12(a): under what circumstances it would be considered “reasonable or appropriate not to take steps or measures”
- Section 12(b): the ambit of the steps required to establish “reasonably practicable”

We recommend that these ambiguities be cleared up with (i) prescribed tests to establish compliance which are clear and precise, and/or (i) specification of the types of actions or steps required to establish compliance.

12. Please explain your reasons.

Harm is caused not just by communicators but also occurs because administrators and platforms play a part in causing or enabling such harm to be committed. All three entities need to act responsibly in order not to cause harm or enable harm to be caused. Therefore we agree that all three entities have to be legally obligated to not cause harm and/or to minimise harm which is committed by them or through their online location or platform.

Please also see the answer to question 11.

B3: Types of relief available from the Courts

If the Court finds that a party has breached their duty under the statutory torts, the Court may order that party to pay damages to the victim. Also, the Court may make orders that the harm be stopped (e.g. be taken down by the communicator) or be further prevented (e.g. a platform may be ordered to shut down an account which has been repeatedly used to harass other users). This is consistent with existing practices for court claims under general tort law.

13. Do you agree that victims of online harms should be able to claim damages in Court?

Yes

14. Do you have any other views on the proposed reliefs which the Court can grant?

We suggest that damages be considered broadly to include restitution on the basis of any unjust enrichment that the perpetrator of harm may have obtained. Perpetrators should not only make good the harm caused by way of compensatory damages, they should also not be allowed to benefit from the tort committed. Any such enrichment gained should be paid over to the victim-survivor.

In addition, we suggest that Courts have the power to make specific orders directing perpetrators to take such actions as the Court may deem fit and appropriate to make good or compensate for the harm caused. These should include the types of directions that the Agency can issue to communicators, administrators and platforms.

Section C: Increasing accountability through improved user information disclosure

We are exploring a proposal to make the user information of perpetrators available to victims who have filed complaints with the Agency (see paragraphs 31-32 of detailed consultation paper), while adhering to data protection requirements under the Personal Data Protection Act (PDPA).

15. Do you agree that it would be useful to disclose a perpetrator's user information to the victim for certain specified purposes (e.g. to bring a claim under the statutory torts, safeguard oneself from the perpetrator) to improve accountability and deter anonymous perpetrators of online harms?

Yes.

16. Please explain your reasons, and any other concerns you might have about this proposal.

The anonymity in many instances online has enabled perpetrators of online harm to cause abuse without fear of punishment or other personal consequences. Furthermore, victim-survivors who suffer harm should rightly be entitled to remedies for such harm against the perpetrator of the harm.

For deterrence purposes and to enable victim-survivors to effectively pursue claims for remedies such as damages (including restitutionary damages), injunctions and specific orders (as recommended in the answer to question 14), victim-survivors need to be able to identify the perpetrator in order to be able to sue them. For this reason, we agree with this proposal to disclose a perpetrator's user information to the victim-survivor for the specified purposes.