IPS Working Papers

ONLINE HARMS IN SINGAPORE:
FROM EVIDENCE TO ACTION

CHEW HAN EI
CAROL SOON
and
HARKIRAN KAUR

November 2025
IPS Working Papers No. 68

Institute of Policy Studies

About Institute of Policy Studies (IPS)

The Institute of Policy Studies (IPS) was established in 1988 to promote a greater awareness of policy issues and good governance. Today, IPS is a think-tank within the Lee Kuan Yew School of Public Policy (LKYSPP) at the National University of Singapore. It seeks to cultivate clarity of thought, forward thinking and a big-picture perspective on issues of critical national interest through strategic deliberation and research. It adopts a multi-disciplinary approach in its analysis and takes the long-term view. It studies the attitudes and aspirations of Singaporeans which have an impact on policy development and the relevant areas of diplomacy and international affairs. The Institute bridges and engages the diverse stakeholders through its conferences and seminars, closed-door discussions, publications, and surveys on public perceptions of policy.

IPS Working Papers No. 68

ONLINE HARMS IN SINGAPORE: FROM EVIDENCE TO ACTION

Chew Han Ei
Senior Research Fellow
Head, Governance & Economy
Institute of Policy Studies
National University of Singapore
han.chew@nus.edu.sg

Carol Soon
Associate Professor (Practice)
Deputy Head, Department of Communications and New Media
National University of Singapore
carol.soon@nus.edu.sg

and

Harkiran Kaur
Research Assistant
Institute of Policy Studies
National University of Singapore
harkiran@nus.edu.sg

In collaboration with Professor Lim Sun Sun (Singapore Management University), Dr Natalie Pang (National University of Singapore) and Natalie Chia (SG Her Empowerment)

November 2025

CONTENTS

ACKNOWLEDGEMENT OF RESEARCH COLLABORATORS

We would like to express our sincere gratitude to Dr Natalie Pang, Dr Lim Sun Sun and Natalie Chia for their invaluable collaboration throughout this project. Their thorough review and constructive feedback enhanced the quality and clarity of our work.

We extend a special thanks to Natalie Chia from SG Her Empowerment for her crucial role in conducting the in-depth interviews with victims and supporters. Her dedication and sensitivity in this process were instrumental in ensuring that these voices were heard and meaningfully represented.

ONLINE HARMS IN SINGAPORE: FROM EVIDENCE TO ACTION

EXECUTIVE SUMMARY

Digital technologies have become an indispensable part of modern life, providing opportunities for connection, information and innovation. The rapid adoption of digital technologies has inadvertently given rise to an array of online harms that pose significant challenges to individuals, communities and societies worldwide.

These challenges are increasingly prevalent in Singapore. While existing legislations address egregious content, there remains no clear legal or shared definition of online harms. This gap allows content which may not meet the legal thresholds for take down to continue circulating online, bringing about deleterious consequences to vulnerable groups.

At the same time, users often struggle to access meaningful recourse. Many are unaware of available support or feel that taking action would have little effect. This lack of clarity and confidence can fuel public apathy, allowing harmful behaviours to become normalised.

This research study, funded by the Ministry of Digital Development and Information, seeks to:

1. Establish a ground-up understanding about definitions and severity of online harms within Singapore's socio-cultural context
2. Understand Singaporeans' responses to online harms and perceptions about effective measures to tackle online harms
3. Formulate recommendations to support the government's efforts to enhance online safety for all Singaporeans

To explore these, we embarked on a three-phased study comprising (1) a landscape scan, (2) focus group discussions (FGDs) and in-depth interviews (IDIs) and (3) a validation survey.

Developing a Localised Online Harms Typology

In Phase One of our study, we reviewed online safety regulations, social media community guidelines and academic literature to develop a localised typology of online harms. Through this review, we identified 20 online harms covered by both regulations and the community guidelines of social media platforms. Our review also revealed harms there are covered only by platforms or regulations.

For the FGDs (Phase Two), we shortlisted 12 online harms based on the localised typology, which participants were asked to rank in terms of severity. Across participants, child sexual exploitation and abuse, violent or terrorism-

related content, and content supporting vice and organised crime were consistently identified as the top three most severe online harms.

Validating the Typology and Understanding Public Perceptions

In Phase Three, we conducted a survey to statistically validate the findings from the preceding phases.

Perceived severity of online harms

Severity perceptions varied across the different online harm types covered, with some consistently seen as more serious and others viewed as less concerning. The three online harms most consistently rated as severe were non-consensual sexual content, the promotion of dangerous behaviours, and targeted harassment. In contrast, harms such as false statements about individuals, hate speech directed at groups outside of Singapore, and online statements instigating disproportionate harm were viewed as less severe.

Severity perceptions were influenced by contextual factors such as consent, timing, geographic proximity, frequency of exposure and the perceived severity of impact.

Factors that influence perceptions of online harms

Singaporeans' perceptions of online harms were primarily shaped through a victim-centric lens. The most influential consideration was the extent of harm to individuals, closely followed by the perceived vulnerability of potential victims.

Stakeholders' responsibility to improve online safety

Improving online safety was seen as a shared obligation, with strong consensus that all stakeholder groups, including users themselves, should do more. Expectations were higher for institutional actors such as the Singapore government, tech companies including social media and app distributors, and individual users. Respondents had lower expectations for social service agencies and non-governmental organisations, though about two-thirds still thought that these groups should do "much more" or "a great deal more."

Perceived usefulness of online harms remedies

The two most useful remedies were holding perpetrators accountable through legislation and ensuring that social media services remove harmful content and accounts more quickly. Other legal remedies — such as the removal of harmful content, empowering victims to seek relief and financial damages, and public education on the existing legal protections — were rated as "very helpful" or "extremely helpful" by more than 70 per cent of respondents.

Recommendations

Drawing from insights across all three study phases, we propose the following recommendations to address online harms.

1. Reinforce What's Working

Enhance victim support pathways

The forthcoming Online Safety Commission should prioritise demystifying the reporting process and clarifying the institutional roles of different stakeholders. It should also develop accessible, multilingual resources that clearly outline steps for filing complaints, expected timelines, and available support services.

Adapt legislation to address emerging online harms

Legal reforms should explicitly criminalise deepfake pornography, with clauses that do not require proof of intent or awareness of harm, given the scalability of generative tools.

Tailor and target digital literacy initiatives better

Harm-specific education should be developed and tailored by age and gender for schools, workplaces, and community settings. Programmes should also offer actionable, tool-based tips for parents and educators, and involve industry and social service agencies in co-developing guides for when harms occur.

Strengthen collaborations within the ecosystem

Singapore should institutionalise joint working groups with policymakers, platforms, and non-governmental organisations to keep protocols and legislation current. This collaborative model reinforces Singapore's regional leadership and strengthens defences against cross-border harms.

2. Address What's Lacking

Address ambiguities to improve reporting and enforcement, and correct misconceptions

Policymakers and platforms should publish clear, regularly updated reporting guidelines and use plain language to explain legislative provisions. These efforts should be supported by public education and outreach through community networks to reach less digitally literate groups.

Improve platform responsiveness and transparency

Platforms should publish regular reports on response times, outcomes, and interventions, and minimum standards should be mandated by the authorities. Safety-by-design features like friction prompts and youth-specific settings should be embedded, alongside joint training for frontline responders, educators, and content moderators.

Create a unified national platform for sharing best practices

A national platform should be developed to share anonymised cases, toolkits, legal updates, and educational resources. This should be a collective endeavour supported by regulators, industry and social service agencies to improve coordination and foster innovation.

Build a culture of shared responsibility

Public campaigns must challenge attitudes that trivialise harm, using personal stories to humanise its impact. Youth should co-design interventions, and

outreach should extend to workplaces using proven models that build awareness and response capability.

3. Anticipate What's Emerging

Offender-centric approaches — Rehabilitation and restorative justice

Diversion programmes should focus on education, empathy and reconciliation, especially when harm stems from ignorance. These must be carefully evaluated for appropriateness, managed to prioritise victim needs, and supported by schools, social services, and the tech industry.

Cross-border online harms

Singapore should expand bilateral and multilateral agreements to improve investigation and enforcement of cross-border offences. Initial efforts can focus on clearly defined areas like scams, child sexual abuse content and non-consensual sexual imagery, using existing regional platforms to build coordination and exert collective pressure on platforms to respond to data requests and content takedowns.

ONLINE HARMS IN SINGAPORE: FROM EVIDENCE TO ACTION

1. INTRODUCTION

The evolution of digital technologies has redefined the way people connect, communicate and seek information. A survey by We Are Social and Meltwater (2024) found that the primary motivations for using the internet are information search, maintaining social connections and entertainment. Beyond these uses, the internet and social media have increasingly become a platform for advocacy, empowering individuals — especially marginalised communities — to raise awareness on social issues, such as the Black Lives Matter movement (Bestvater et al., 2023).

The COVID-19 pandemic played a pivotal role in accelerating global reliance on digital platforms. As countries imposed lockdowns and social distancing measures, people increasingly turned to digital platforms for work (Hern, 2020), education (European Training Foundation, 2020), social connection (Szeto et al., 2024), and access to essential services (Organisation for Economic Co-operation and Development, 2021). This unprecedented shift dramatically increased the time individuals spent online, inadvertently heightening their vulnerability to an array of online harms. In Australia, for example, reports of illegal content rose by 90 per cent in 2020, with sharp increases in image-based abuse, online harassment, and cyberbullying (Grant, 2021).

1.1. Online Harms Landscape in Singapore

With one of the highest internet penetration rates worldwide (Infocomm Media Development Authority, 2023a), Singapore is not immune from online harms. A national survey by the Ministry of Digital Development and Information (2024) found that two-thirds of Singapore users encountered harmful content online.

Despite the increasing encounters with online harms, Singapore, like many countries, lacks a clear legal and community-based definition of online harms. While the Online Safety (Miscellaneous Amendments) Act 2022 identifies "egregious content" — such as posts that promote self-harm, suicide, child sexual exploitation and abuse, terrorism and racial or religious tensions, or pose a risk to public health — there remains a significant gap in addressing less extreme but still harmful online behaviours.

A commentary written by the first and second author of this working paper (Chew et al., 2022) highlighted that this legal and conceptual gap allows potentially damaging content such as harmful social media challenges to circulate under the radar, leaving vulnerable users at risk. This disconnect can also foster public apathy, where such actions are perceived as harmless.

The authors also drew attention to the language used to describe online harms, noting that it could lead to unintended implications. For example, terms like "revenge porn" may carry the connotation of "victim blaming" while "sexting" may trivialise the impact on victims. Furthermore, many of these terms originate

from foreign contexts and might not reflect Singapore's unique sociocultural context (Chew et al., 2022).

Beyond definitional clarity, victim support systems are another area with significant gaps. A survey conducted by the Sunlight Alliance for Action found that 57 per cent of respondents were unaware of help-seeking avenues, while 43 per cent who had experienced online harms chose not to act, believing it would not make a difference (Ministry of Digital Development and Information, 2022). This is echoed by a study conducted by the second author, which found that most Singaporeans tended to ignore false information they encountered online, likely due to apathy or a lack of knowledge on available recourse (Soon et al., 2023). There is a clear operational gap in supporting victims to seek recourse and receive assistance when they experience online harms.

This research study, funded by the Ministry of Digital Development and Information, seeks to:

1. Establish a ground-up understanding about definitions and severity of online harms within Singapore's sociocultural context
2. Understand Singaporeans' responses to online harms and perceptions about effective measures to tackle online harms
3. Formulate recommendations to support the government's efforts to enhance online safety for all Singaporeans

To determine Singaporeans' perceptions of online harms, their thresholds for determining harm, and their views on the effectiveness of existing legislations, platform policies, and education efforts, we conducted a study over three phases.

In Phase One, we conducted a review of regulations, social media community guidelines and academic literature to better understand various online harms, their impacts on individuals and communities, and the available legislative remedies and platform measures to tackle online harms. See Appendix A, Tables A1 and A2 for a summary of the regulations and community guidelines reviewed.

In Phase Two, we conducted focus discussion groups (FGDs) with 79 participants, including youths (aged 17 to 35 years old), parents, educators and social service professionals, industry partners, and safety tech providers. The discussions allowed us to examine severity perceptions, factors influencing severity perceptions, public awareness of existing policies, safeguards, and legislations related to online safety. It also allowed us to evaluate public sentiments of current policies and identify areas for improvement. See Appendices C and D for more details on the methodology used.

Concurrently, in-depth interviews (IDIs) with victims and supporters were conducted by our research collaborator, SG Her Empowerment. Through these interviews, we obtained first-hand accounts of victims' experiences as well as the barriers they encountered to reporting online harms incidents, the

effectiveness of available support structures, and their coping mechanisms (see Appendix E).

In Phase Three, we conducted a validation survey to statistically validate the findings obtained from the preceding phases (see Appendix F).

2. DEVELOPING A LOCALISED TYPOLOGY OF ONLINE HARMS

In 2023, the World Economic Forum (WEF) published an online harms typology which proposes a common lexicon for online harms that affect individuals and society (World Economic Forum, 2023). The typology groups 25 online harms into six broad categories: (1) threats to personal and community safety, (2) harm to health and well-being, (3) hate and discrimination, (4) violation of dignity, (5) invasion of privacy and (6) deception and manipulation.

The WEF typology also identifies the type of risk a particular harm would engender. The three types of risk are: (1) content risks, (2) contact risks and (3) conduct risks (see Table 1). Although the typology addresses a broad range of online harms, WEF acknowledges that local nuances and sociocultural factors may necessitate adaptations for regional relevance (World Economic Forum, 2023). Phase One of the study builds on the WEF's work to develop a Singapore-specific typology of online harms that reflects local realities and priorities.

Table 1: Summary of the WEF online harms typology

Type	Content risks	Contact risks	Conduct risks
Threats to personal and community safety	1. Child sexual abuse material 2. Child sexual exploitation material 3. Pro-terror material 4. Content that praises, promotes, glorifies or supports extremist organisations or individuals 5. Violent graphic content 6. Content that incites, promotes or facilitates violence 7. Content that incites, promotes or instructs dangerous physical behaviour	1. Grooming for sexual abuse 2. Recruitment and radicalisation	1. Technology-facilitated abuse 2. Technology-facilitated gender-based violence
Harm to health and wellbeing	1. Material that promotes suicide, self-harm and disordered eating 2. Developmentally inappropriate content		
Hate and discrimination	1. Hate speech		1. Algorithmic discrimination
Violation of dignity		1. Sexual extortion	1. Online bullying and harassment
Invasion of privacy			1. Doxxing 2. Image-based abuse
Deception and manipulation	1. Disinformation and misinformation 2. Deceptive synthetic media		1. Impersonation 2. Scams 3. Phishing 4. Catfishing

Adapted from: World Economic Forum. (2023). *Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms*. https://www3.weforum.org/docs/WEF_Typology_of_Online_Harms_2023.pdf

In July 2023, the Infocomm Media Development Authority (IMDA) introduced the Code of Practice for Online Safety – Social Media Services (CoP), requiring designated social media platforms to curb harmful content (Infocomm Media Development Authority, 2023b). The CoP addresses six categories of online harms: (1) sexual content, (2) violent content, (3) suicide and self-harm content, (4) cyberbullying content, (5) content endangering public health and (6) content facilitating vice and organised crime.

Through our review of online safety regulations and community guidelines, we found nine harms that were not included in both the WEF typology and IMDA's CoP (see Appendix Table A3). Additionally, there are harms that are closely related and often described together in regulations and community guidelines. For parsimony, we described online harms that are broadly similar across the WEF typology, IMDA's CoP, other regulations and community guidelines together. By assessing the various legislations and guidelines, we identified gaps in coverage of online harms and laid the groundwork for the development of a localised online harms typology that addresses local context and needs (see Appendix B for coverage of online harms).

2.1. Testing the Typology: Ranking Harms in a Local Context

For the FGDs, we adapted the online harms typology from Phase One for a ranking activity. Harms that were specifically covered by the IMDA CoP and Singapore Online Criminal Harms Act 2023 formed the core list for participants to evaluate (see Appendix D).

Meanwhile, harms such as scams and cybercrimes were excluded from the ranking activity. Given their prominence in public discourse and widespread recognition as serious threats in Singapore, including them would have skewed the rankings of other lesser-known harms.

Other harms identified during the landscape scan — such as human trafficking or sexual solicitation — were either synthesised into broader categories (e.g., content supporting vice and organised crime) or included in a supplementary list.

Findings from our landscape scan and FGDs consistently identified child sexual exploitation and abuse as the most severe online harm. It is a central focus in most of the regulations and guidelines reviewed, with 27 out of 42 regulations and guidelines covering the harm (see Appendix B). Additionally, an overwhelming majority of our focus groups participants (67 out of 79) ranked it as the most severe harm, placing it first, second or third.

> "Morally, a lot of people would think [child sexual exploitation and abuse] is unacceptable. I believe it should be the most severe, if not, one of the most severe online harms." **— P409, Male, 31 years old, User**

> "I put child sexual exploitation and abuse as most severe because there is always a market for it… the reality is that children are extremely vulnerable… which makes it easier for them to be exploited." **— P801, Female, 25 years old, Social Service Professional**

This is followed by violent or terrorism-related content. Our landscape scan highlighted that such content is among the most heavily regulated online harms, with 30 out of 42 regulations and guidelines covering content portraying graphic violence, and 26 out of 42 regulations and guidelines covering content promoting or supporting terrorist or extremist organisations (see Appendix B). Similarly, 50 out of 79 FGD participants assigned the harm as one of the top three most severe harms, citing its potential to cause societal harm.

> "I rank violent or terrorism-related content as the most severe because it has the potential to cause the greatest harm to society. In my view, any content that normalises violence is the most severe." **— P602, Male, 36 years old, User**

> "Terrorist acts can cause a lot of deaths and injuries… that's why I ranked it as the most [severe]." **— P503, Female, 39 years old, User**

Content supporting vice and organised crimes was ranked as the third most severe harm. Harms such as content facilitating or promoting the sale of regulated, restricted or illegal goods and services, cheating, fraud, scams, sexual solicitation, human trafficking, and promotion of illegal activities or behaviours are covered by at least 15 out of 42 regulations and guidelines reviewed (see Appendix B). Similarly, 24 out of 79 FGD participants ranked the harm as one of the top three most severe harms.

> "Vice and organised crime is high on my list because I find that criminal rings are more likely to have far greater resources to outstrip [sic] a layperson." **— P1005, Female, Safety Tech Provider**

> "The reason why [content supporting vice and organised crime] is not least severe for me is because stuff like scams and gambling do have the potential to destroy families." **— P304, Male, 21 years old, User**

Given the consistency in how child sexual exploitation and abuse, violent or terrorism-related content, and content supporting vice were perceived across both the landscape scan and FGDs, we excluded these harms from Phase Three to obtain more discriminatory results for the remaining harms.

3. ASSESSING PERCEIVED SEVERITY OF ONLINE HARMS

To assess the perceived severity of different types of online harms, we employed a MaxDiff (Maximum Difference Scaling) design (see Appendix G for a detailed explanation). This method allowed us to obtain greater distinction

between the different harms we were testing compared to traditional Likert scale or ranking questions.

The validation survey was conducted in March 2025 with 600 respondents who were roughly representative of Singapore's resident population by gender and age groups. See Figures F1 and F2 in Appendix F for the breakdown of the respondents by age and gender, respectively. A total of 16 online harms were evaluated in this phase. These harms were chosen based on the insights obtained from Phase Two. See Appendix H for list of harms evaluated.

3.1. Rankings of Online Harms

To understand how respondents perceived the severity of online harms, we looked at the difference between how often each harm was selected as the most severe and how often it was selected as the least severe. The greater the difference, the more severe the harm was perceived relative to others (see Figure 1).

Sexual content depicting voyeuristic or intimate images recorded without consent was overwhelmingly perceived as the most severe harm. It was selected 801 times as the most severe and 223 times as the least severe, resulting in a relative score of 578. This was consistent with findings from the landscape scan, where 29 out of 42 regulations and community guidelines addressed such content, indicating its broad recognition as an egregious harm

(see Appendix B). An educator from the FGDs explains why he ranked sexual content as one of the top three most severe harms.

> "It takes just one Telegram group, one message, for my students to be exposed to pornography. And [sexual content] is insidious precisely because it tells them, this is what sex looks like, this is what relationships look like. And for me, it's a language that they use when they talk about other people, about girls. And many cases that I dealt with involve boys. They sort of cross the boundary where they don't know how to talk about girls in a healthy way." **— P808, Male, 45 years old, Educator**

Promotion of dangerous behaviours followed next, receiving a relative score of 335. This harm was also prominently covered in our landscape scan, with more than half of the regulations and community guidelines covering the harm (see Appendix B). FGD participants, especially parents, educators and social service professionals, expressed strong concerns about the emotional impact of such content and its potential to normalise such behaviours.

> "Promotion of dangerous behaviours is one of the most severe harms, mainly because of the potential impact it can have on my daughter's well-being. I don't want her to hurt herself… or be driven to suicide because she feels inadequate." **— P703, Male, 44 years old, Parent**

> "My students often look for content relating to self-harm, depression, anxiety and other mental health issues. They end up normalising these

> behaviours. They will tell themselves, 'I cut myself is normal [sic], so many people are posting about it online.' They will also share different methods for self-harm and suicide." **— P801, Female, 25 years old, Social Service Professional**

Targeted harassment[1] was also considered highly severe, with a score of 292. Cyberbullying or harassment was commonly covered in our landscape scan, with 30 out of 42 regulations and community guidelines directly or indirectly covering the harm (see Appendix B). Both focus group and interview participants highlighted the significant psychological effects of these harms, including loss of self-worth and severe distress. This was especially evident in the case of QV02, who endured persistent cyberbullying and harassment over a period of two years. She described how the continuous abuse — including anonymous phone calls, cyberstalking, and a barrage of derogatory messages from anonymous accounts — began to erode her self-assurance.

> "I mean, did I set this [cyberbullying and harassment] on myself?... Do I do this when I'm having my split [sic]? Do I have split personality disorder?... Maybe I've been awake all this while... when you have been harassed like this... victims of stalking start feeling a loss, a sense of disorientation…You do not know anywhere anymore." **— QV02, Female, 40 years old**

[1] Participants from the FGDs made a distinction between cyberbullying and targeted harassment which typically involves repeated threats that could have long-lasting impacts.

Other harms perceived as severe included sexual content that was consensually recorded, which had 599 most severe and 344 least severe selections, resulting in a relative severity score of 255.

Harms perceived to be of moderate severity included doxxing, which received a relative severity score of 180, followed by impersonation with a score of 94, and content endangering public health with a score of 56. Cyberbullying was also perceived as less severe, receiving a relative of –40.

Further down the scale, the perceived severity of harms declines notably. Both statements affecting a person's reputation and hate speech or slurs against target groups in Singapore had a relative score of –149. Misinformation about health risks ranked even lower, with a relative score of –243.

At the bottom of the rankings were the harms which were perceived to be the least severe. Misuse of inauthentic material had a relative score of –291, while false statements about individuals scored –321. Hate speech or slurs targeting groups outside of Singapore ranked even lower, with a score of –404.

Online statements instigating disproportionate harm was ranked as the least severe harm, with only 221 most severe selections compared to 696 least severe selections, and a relative score of –475.

Survey results aligned with findings from the FGDs regarding the least severe harms. Although online falsehoods and their impact have generated concerns

in recent years, the FGD participants perceived it to be one of the least severe harms, perhaps in part due to the Singapore government's swift actions in curbing them through the Protection from Online Falsehoods and Manipulation Act 2019. Another reason could be that online falsehoods have been normalised as an inherent aspect of the online experience.

> "Out in the online space, there's so much false and misleading material. It's a given that people who use the online space have to contend with." **— P906, Male, Industry Partner**

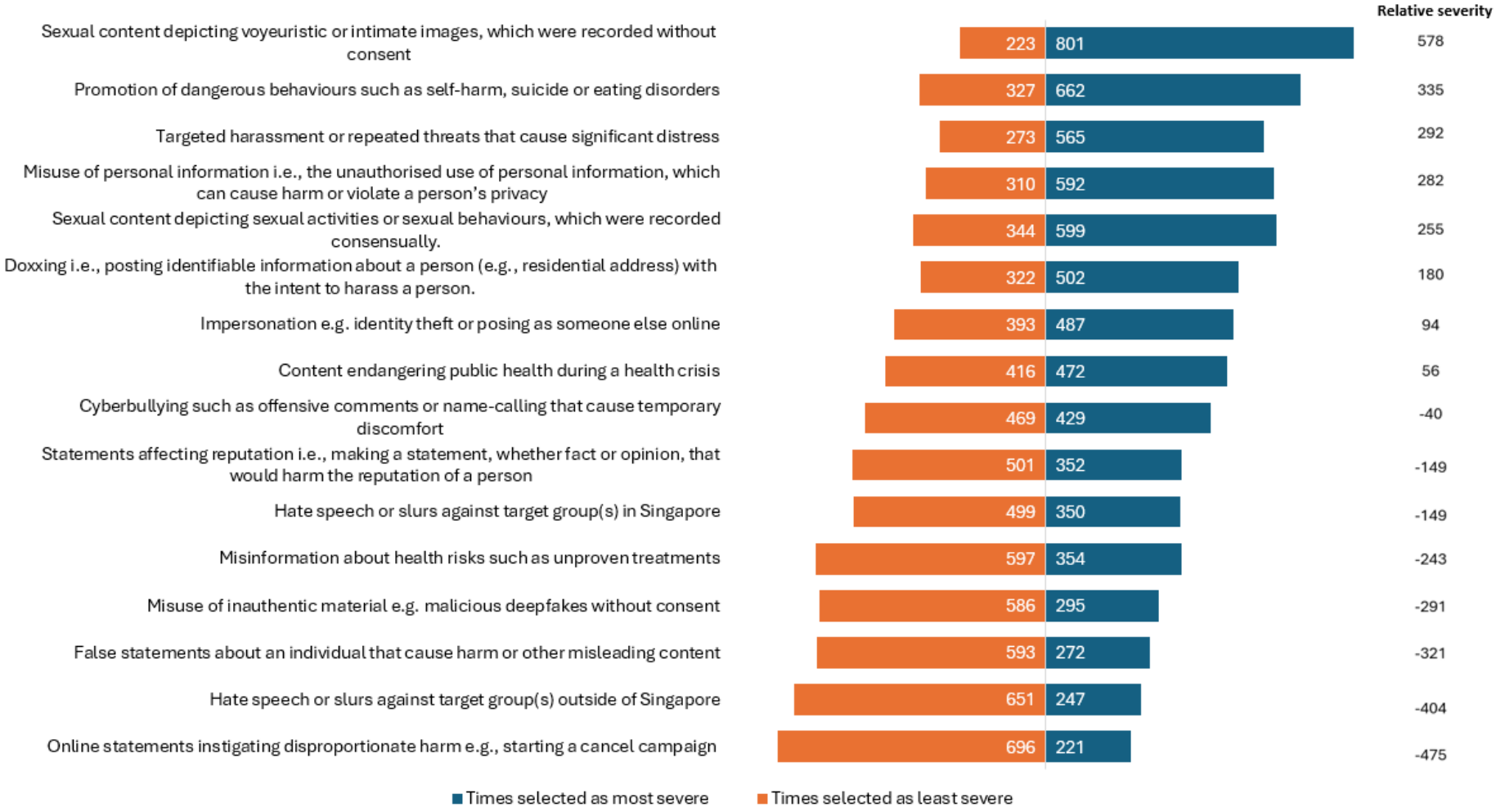
> "It has become so normalised, it has lost some of its impact [sic]. For example, misuse of information or false or misleading information, or like deepfakes, things like that… but there's always this doubt there." **— P506, Female, 46 years old, User**

A similar view was expressed regarding cancel campaigns. While participants recognised cancel campaigns as harmful, it was not prioritised in severity when compared to other online harms.

> "Compared to causing somebody to stab someone, versus causing reputation harm to an individual, I think it's quite clear that this is not as important." **— P602, Male, 36 years old, User**

"I put it lower because it says 'cancel culture,' for me I think [cancel culture is] something stupid." **— P801, Female, 25 years old, Social Service Professional**

Figure 1: Overall rankings of online harms



3.2. Demographic Differences in Severity Perceptions

Our findings showed that Singaporeans' severity perceptions can vary by gender and age groups. Some harms showed broad agreement in perceived severity across different groups, while others revealed moderate to significant variations (see Figure 2).

Overall, sexual content depicting voyeuristic or intimate images recorded without consent was unanimously regarded as the most severe online harm (rank 1). This reflected consensus on the seriousness of privacy and consent violations in digital spaces. Promotion of dangerous behaviours was ranked second across most groups, except females and older respondents who perceived the harm as slightly less severe and ranked it third.

When it comes to moderately severe harms, cyberbullying was ranked at rank ninth (out of 16) across most groups. However, youths perceived the harm as slightly more severe, ranking it at seventh.

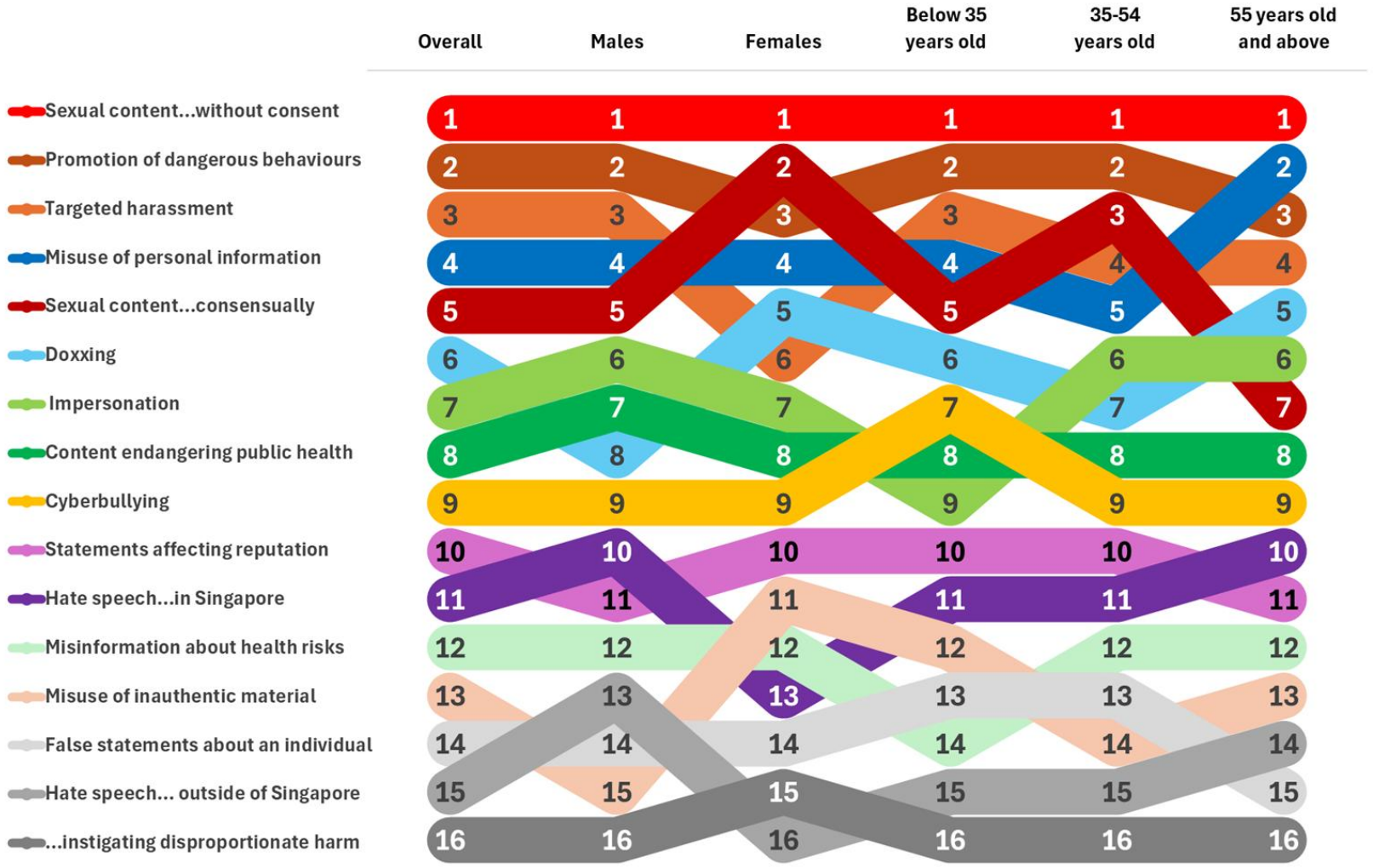
Cancel campaigns were perceived as the least severe harm, ranking at 16th across most groups, except females who perceived the harms as slightly more severe, ranking at rank 15th.

Some harms exhibited greater variation in rankings across demographic groups. For example, sexual content recorded consensually drew mixed views. It was ranked fifth overall but females and middle-aged respondents perceived

it as more severe, placing it at second and third, respectively. Older respondents ranked it lower, at seventh, indicating generational differences in sensitivity to such content.

Larger variation in perceived severity was observed for misuse of inauthentic materials, which was ranked 13th overall. Males and middle-aged respondents perceived the harm as less severe, placing it at 15th and 14th, respectively. In contrast, females and youths perceived the harm to be more concerning, placing it at 11th and rank 12th, respectively. These differences suggest growing concerns about the potential risks posed by generative AI.

Figure 2: Severity rankings, by total sample, gender and age categories



3.2.1. How males and females perceive online harms

Both male and female respondents ranked sexual content depicting voyeuristic or intimate images recorded without consent as the most severe online harm (ranked first). This indicates a universal recognition of the severity of such harm. However, based on relative scores, female respondents tended to perceive this harm more severely than males (females: 369, males: 209). This may reflect the reality that females are disproportionately targeted by non-consensual image-sharing and sexual harms (SG Her Empowerment, 2023). Figures I1 and I2 in Appendix I presents the relative scores for each harm as reported by male and female respondents respectively.

Females were particularly concerned about harms that carry reputational consequences. For example, while consensual sexual content was ranked fifth overall, females ranked it much higher, at second place. This likely relates to sexual double standards that tend to stigmatise female sexuality more harshly (Park et al., 2023), contributing to greater feelings of vulnerability among females.

Personal safety was another key concern for females. This is reflected in their higher ranking of doxxing compared to the overall sample. Research shows that females generally have greater fears about personal safety (Powell et al., 2022). Since doxxing can lead to real-world threats, female respondents might have been more cognisant of such harms compared to males.

Males, on the other hand, placed more emphasis on harms that they are frequently exposed to and those with broader societal implications. For example, males ranked hate speech targeting groups in Singapore (ranked 10th) and outside of Singapore (ranked 13th) higher than females. Research conducted in other contexts sheds light on the reason — males experience hate speech more often (Stevens et al., 2024) which may have influenced their severity perceptions. Additionally, males also ranked misinformation more severely than females in terms of relative scores (females: –147, males: –96). Further research could examine males' online experiences, particularly given the rise of subcultures like manosphere (Venkataramakrishnan & Squirrell, 2024).

Interestingly, males also ranked cyberbullying as more severe compared to females in terms of relative scores (females: –40, males: 0). This could be attributed to the way males are socialised to respond to conflict through confrontation, leading them to perceive name-calling or insults as a direct challenge to their masculinity (Saucier et al., 2025). Both the overall sample and males ranked targeted harassment at third; however females ranked it lower (ranked sixth). This lower ranking does not necessarily indicate that females perceived targeted harassment as unimportant or benign. Rather, it suggests that other types of harms such as sexual content recorded consensually might have resonated more strongly due to their immediate relevance, leading to higher prioritisation in the rankings.

3.2.2. How different age groups perceive online harms

Across all three age groups, sexual content depicting voyeuristic or intimate images recorded without consent was consistently ranked as the most severe harm (rank 1), aligning with the overall sample. While respondents recognised its severity, those between 35–54 years old perceived the harm to be more severe compared to the other age groups (below 35 years old: 177; 35–54 years old: 252; 55 years old and above: 149). Figures I3 to I5 in Appendix I show the relative scores for each harm as reported by each age groups (below 35 years old, 35 to 54 years old and 55 years old and above).

Sexual content recorded consensually was ranked fifth by youths and the total sample. However, middle-aged respondents perceived the harm to be more severe, ranking it at third. In contrast, older respondents viewed the harm as less severe, ranking it at seventh. These differences were consistent with findings from the FGDs, where generational differences influenced what is considered as socially acceptable in relation to nudity and explicit content.

> "Sometimes, it's a negotiation across different generations. For example, the youth will say 'What's wrong if I record myself naked?' We might feel differently because we are looking at it from a different norm [sic]." **— P1002, Male, Safety Tech Provider**

> "I receive many reports from members of public that say, 'Look at this, this is nudity; it is cleavage.' So then, that is a question of interpretation. Do we then take action based on someone's perception or someone taking personal offense?" **— P1005, Female, Safety Tech Provider**

Targeted harassment was ranked third overall and by youths. Respondents aged 35–54 years old and 55 years old and above ranked the harm slightly lower, at fourth. In terms of relative scores, middle-aged respondents perceived the harm to be more severe compared to older respondents (35–54 years old: 120; 55 years old and above: 70).

Respondents aged 35–54 years old and 55 years old and above ranked cyberbullying at rank 9, consistent with the overall sample. Youths, on the other hand, perceived the harm to be more severe, ranking it two ranks higher, at seventh. These findings aligned with those from the FGDs where older participants expressed greater resilience to harms such as cyberbullying, citing reduced online engagement and being less affected by peer speech or actions. Another explanation for how cyberbullying was ranked by people from different age groups is the frequency of exposure to cyberbullying. Youths tend to experience cyberbullying more frequently than older age groups (Wang et al., 2019), which likely influenced their severity perceptions.

In contrast, youths ranked misinformation as less severe, possibly due to their higher digital literacy and greater self-confidence in identifying false content (Gottfried &

Grieco, 2018), making it feel less personally threatening compared to other online harms.

Content endangering public health was ranked eighth across the age groups, matching the broader sample. However, older respondents perceived the harm to be more severe compared to other age groups (below 35 years old: 0; 35–54 years old: 26; 55 years old and above: 30). On the other hand, respondents aged 35–54 years old and 55 years old and above ranked misinformation about health risks at 12th, similar to the broader sample. In terms of relative scores, older respondents perceived the harm to be more severe compared to middle-aged respondents (35–54 years old: –84; 55 years old and above: –62). In contrast, youths perceived the harm to be less severe, ranking it two ranks below, at 14th. This points to a generational difference in how misinformation regarding health matters is perceived. In general, health information is important to older people compared to the young.

Impersonation was ranked seventh overall but middle-aged respondents and older respondents perceived the harm to be slightly more severe, ranking it at sixth. On the other hand, youths perceived the harm to be less severe, placing it at ninth. This could be because as digital natives, younger users are more familiar with impersonation-like behaviours like parody and fan accounts (Baltezarevic & Baltezarevic, 2024) and often view them as harmless or part of normal online culture. This normalisation may have suppressed youths' severity perceptions of impersonation. On the other hand, older respondents may be more sensitive to impersonation due to the perceived proximity to scams and other cybercrime threats.

Hate speech against target groups in Singapore was ranked 11th overall and by youths and middle-aged respondents, with youths perceiving it to be more severe compared to middle-aged respondents (below 35 years old: –56; 35–54 years old: –72). In contrast, older respondents perceived the harm to be more severe, ranking it slightly higher, at 10th. Regarding hate speech against target groups outside of Singapore, younger respondents ranked the harm at 15th, consistent with the larger sample, with youths perceiving it to be more severe compared to middle-aged respondents (below 35 years old: –125; 35–54 years old: –164). However, older respondents perceived the harm to be more severe, placing it at 14th. The concerns older respondents have over hate speech in general could stem from their experiences or greater familiarity with the importance of inter-community relationships on society.

Misuse of inauthentic material was ranked 13th overall and by older respondents. Interestingly, youths perceived the harm as more severe, placing it at 11th, while middle-aged respondents perceived the harm to be less severe, ranking it at 14th. This suggested that youths may be more attuned to emerging digital threats, likely due to their greater exposure and engagement with technology and online content manipulation.

3.3. Drawing the line on online harms

3.3.1. Consent

In Phase Two of the study, FGD participants pointed out that the category of sexual content was overly broad. They emphasised that specific forms of harm, such as

sextortion and non-consensual image distribution, are more severe than consensual depictions of sexual activity.

> "When we talked about voyeuristic or intimate imagery, it's not clear if you're talking about the perpetrator and what kind of content we're talking about. There are also overlaps with CSAM. So, it depends on age verification and our ability to identify [the content type]. And then you take explicit sexual activities, that's what we call pornography. Even then, it's people's personal freedom if they want to watch that stuff. There are certain things like deviant sexual behaviour, voyeuristic imagery, and sextortion, those are a lot more severe. So, in my personal view, if you take sexual activities outside of that [category] and you leave in the rest, I would definitely think of that [category] as a lot more severe." **— P901, Female, Industry Partner**

> "I feel there should be a clear distinction between non-consensual and consensual sexual content, as the two are drastically different." **— P402, Male, 26 years old, User**

Similarly, our landscape scan revealed that social media platforms often ban sexual content that is gratifying, distressing, fetishising or fictional — such as in YouTube's policy on Nudity and Sexual Content (Google, n.d.). However, some platforms make exceptions. For instance, LinkedIn's policy on adult nudity and sexual content states that while such content is generally removed, certain instances may be permitted if relevant to professional discussions (LinkedIn, n.d.).

Compared to social media platforms, regulations typically only address adult nudity and sexual content if it is distressing or extremely graphic. For example, the list of offence-specific categories in Ireland's Online Safety and Media Regulation Act 2022 includes "online content by which a person exposes his or her genitals intending to cause fear, distress or alarm to another person" (p.106) and "online content by which a person intentionally engages in offensive conduct of a sexual nature" (p.107).

Findings from our survey validated these insights, with respondents ranking sexual content recorded without consent as the most severe harm (ranked first) and sexual content recorded consensually ranked fifth. Preference shares[2] supported this (see Table 2), with sexual content recorded without consent receiving the highest preference share of 10.8 per cent compared to 8.4 per cent for sexual content recorded consensually (see Appendix G for more information about preference shares).

3.3.2. Frequency and severe individual impacts

Informed by the FGD findings, we also tested whether severity of impact and frequency of exposure influenced severity perceptions by comparing rankings of targeted harassment and cyberbullying.

Survey respondents perceived targeted harassment (ranked third) to be more severe than cyberbullying (ranked ninth). Preference shares reflected this difference, with

[2] Preference shares from the MaxDiff method provide a model based on the proportion of relative importance that adds up to 100 per cent.

targeted harassment receiving 8.0 per cent compared to 5.9 per cent for cyberbullying (see Table 2). This aligned with FGD whereby educators and social service professionals shared how trolling and bullying can be protracted behaviours that inflict emotional distress on victims over time.

> "Even if you block someone, they can create another account, they can use many other platforms like games. So sometimes, when students know they cannot block someone, they find other ways to handle this, like going into their shell. They also get affected by one other, and they can get very disturbed. [Students] react in different ways. Some engage in behaviours that are not helpful, and others close up [sic]." **— P807, Male, 37 years old, Educator**

> "If a person is bullied or beaten up, they are sent to a hospital. There is a medical report and a timeline for recovery. But with cyberbullying, 'what's the medical report?', 'what does recovery even look like?'… and [cyberbullying] does not stop at one wave, there can be a second wave. There are so many unknowns." **— P803, Male, 28 years old, Social Service Professional**

3.3.3. Time sensitivity

We also examined how time sensitivity affects perceptions of severity. Survey respondents perceived content endangering public health during a health crisis (ranked eighth) to be more severe compared to misinformation about health risks such as unproven treatments (ranked 12th). Preference shares confirmed this, with content

endangering public health during a health crisis receiving 6.6 per cent compared to 4.6 per cent for misinformation about health risks such as unproven treatments (see Table 2).

This aligned with insights from industry stakeholders in our focus groups, who explained that the perceived severity of misinformation increases during crises like the COVID-19 pandemic but regresses to normal during non-crises.

> "You're less worried about electoral misinformation when there's no election going on. You're less worried about information that could endanger public health if there's no public health emergency. But if there is a public health emergency, suddenly that shoots up and then it's much more severe. So, I think specific events or trends can significantly amplify the harms in these kinds of category." **— P905, Male, Industry Partner**

> "In an unprecedented situation, like COVID-19, that's when the severity can change… When you think about platforms which allow people to make anti-vaxxing speech, that's a political stance. We have to preserve an element of free speech as platforms. One might say you need to remove this person and the content from the platform. This is where I won't say I want to remove this person from the platform for having an opinion about vaccinations. But I might want to reduce the virality of this kind of content." **— P901, Female, Industry Partner**

Remedies and responses should take into account the influence of context of time on the severity of an online harm.

3.3.4. Geographic proximity

Finally, validation results showed that geographic proximity influences how harms are perceived. Hate speech or slurs against target groups in Singapore (ranked 10th) was perceived as more severe compared to hate speech or slurs against target groups outside of Singapore (ranked 15th). Preference shares supported this, with 5.4 per cent for hate speech in Singapore compared to 3.7 per cent for hate speech outside Singapore (see Table 2). This is consistent with FGD findings where older female participants shared how harms occurring closer to Singapore was more serious than those taking place in other regions.

> "In the Middle East, maybe [terrorism-related content] will be ranked as most severe." **— P502, Female, 37 years old, User**

> "[Terrorism] doesn't happen here in Singapore, so [the terrorism-related content] doesn't affect Singapore." **— P501, Female, 36 years old, User**

Table 2: Preference shares derived from MaxDiff method

Types of online harms	Preference shares Total: 100%
Sexual content depicting voyeuristic or intimate images, which were **recorded without consent**	10.8
Promotion of dangerous behaviours such as self-harm, suicide or eating disorders	8.9
Sexual content depicting sexual activities or sexual behaviours, which were **recorded consensually**	8.4
Misuse of personal information, i.e., the unauthorised use of personal information, which can cause harm or violate a person's privacy	8.1
Targeted harassment or repeated threats that cause **significant distress**	8.0
Doxxing, i.e., posting identifiable information about a person (e.g., residential address) with the intent to harass a person	7.4
Impersonation, e.g., identity theft or posing as someone else online	6.7
Content endangering public health **during a health crisis**	6.6
Cyberbullying such as offensive comments or name-calling that cause **temporary discomfort**	5.9
Hate speech or slurs against target group(s) **in Singapore**	5.4
Statements affecting reputation, i.e., making a statement, whether fact or opinion, that would harm the reputation of a person	4.8
Misinformation about health risks such as unproven treatments	4.6
Misuse of inauthentic material, e.g., malicious deepfakes without consent	4.0
Hate speech or slurs against target group(s) **outside of Singapore**	3.7
False statements about an individual that cause harm or other misleading content	3.6
Online statements instigating disproportionate harm, e.g., starting a cancel campaign	3.1

Note: Harms that are colour-coded were compared against one another for hypotheses testing.

3.4. How Online Harms Rank Against One Another

A comparison table was also generated using preference shares to calculate head-to-head comparisons between harms (see Table 3). The percentages[3] in the table represent the percentage of the times that the harm listed on the left side of the table was rated as more severe than the corresponding harm across the top of the table.

The table shows consistent patterns in how respondents perceived the relative severity of online harms. Sexual content depicting voyeuristic or intimate images recorded without consent was consistently rated as the most severe harm, outperforming nearly all other harms in pairwise comparisons. It has especially high percentages against online statements instigating disproportionate harm; false statements about an individual; misuse of inauthentic material; misinformation; and hate speech or slurs against target groups outside of Singapore. It was also rated more severe than consensual sexual content 56.3 per cent of the time.

Promotion of dangerous behaviours was also rated as highly severe, frequently surpassing other harms including hate speech or slurs against target groups outside of Singapore, false statements about an individual, and online statements instigating disproportionate harm.

[3] Each value was calculated by dividing the preference share of one harm by the sum of the preference shares of both harms being compared. For example, to estimate the likelihood that sexual content depicting voyeuristic or intimate images, which were recorded without consent would be rated as more severe than online statements instigating disproportionate harm, the relative probability would be calculated using this formula: 10.8 / (10.8 + 3.1) = 77.7. This means that respondents would choose sexual content depicting voyeuristic or intimate images, which were recorded without consent as more severe 78 per cent of the time compared to online statements instigating disproportionate harm.

Targeted harassment was also rated as more severe than online statements instigating disproportionate harm 72.1 per cent of the time, and cyberbullying 57.6 per cent of the time, suggesting that harms with repeated exposure and higher impact was perceived as more damaging.

At the lower end of the severity scale were online statements instigating disproportionate harm and false statements about an individual, both of which received less than 50 per cent in most pairwise comparisons.

The table also allows for all other pairwise comparisons, such as between promotion of dangerous behaviours and misuse of inauthentic material. In this case, promotion of dangerous behaviours would be chosen as more severe 69.0 per cent of the time.

Table 3: Head-to-head comparison

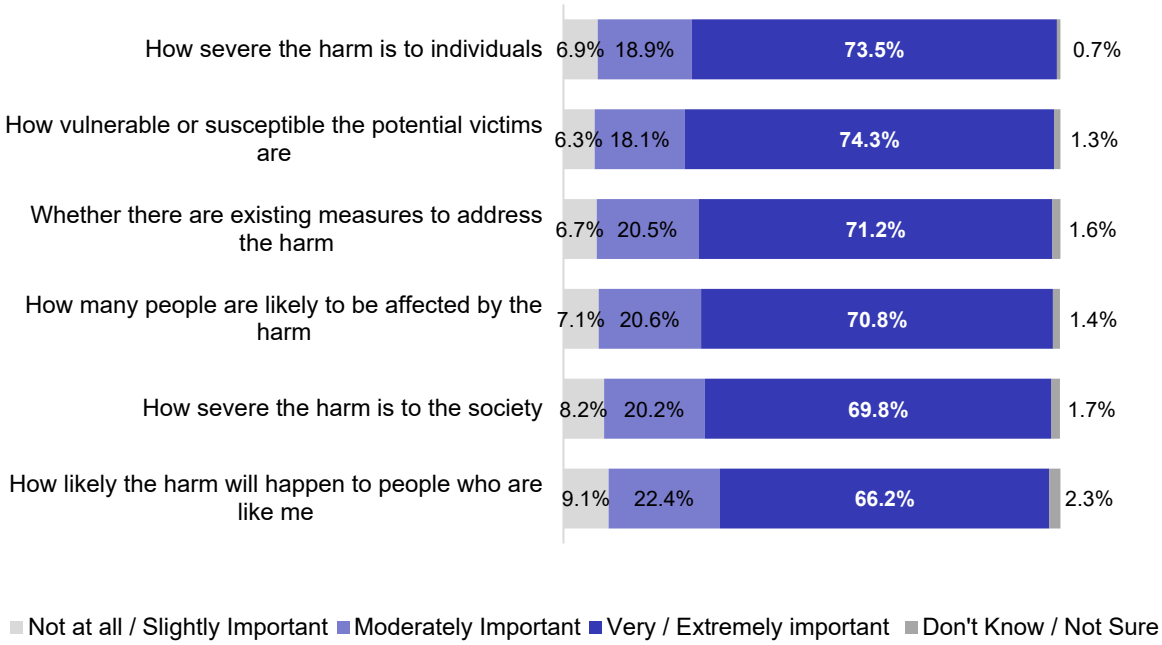
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Sexual content depicting voyeuristic or intimate images, which were recorded without consent		56.3	54.8	64.7	57.4	66.7	74.5	62.1	70.1	73.0	75.0	61.7	77.7	69.2	57.1	59.3
2. Sexual content depicting sexual activities or sexual behaviours, which were recorded consensually.	43.8		48.6	58.7	51.2	60.9	69.4	56.0	64.6	67.7	70.0	55.6	73.0	63.6	50.9	53.2
3. Promotion of dangerous behaviours such as self-harm, suicide or eating disorders	45.2	51.4		60.1	52.7	62.2	70.6	57.4	65.9	69.0	71.2	57.1	74.2	65.0	52.4	54.6
4. Cyberbullying such as offensive comments or name-calling that cause temporary discomfort	35.3	41.3	39.9		42.4	52.2	61.5	47.2	56.2	59.6	62.1	46.8	65.6	55.1	42.1	44.4
5. Targeted harassment or repeated threats that cause significant distress	42.6	48.8	47.3	57.6		59.7	68.4	54.8	63.5	66.7	69.0	54.4	72.1	62.5	49.7	51.9
6. Hate speech or slurs against target group(s) in Singapore	33.3	39.1	37.8	47.8	40.3		59.3	45.0	54.0	57.4	60.0	44.6	63.5	52.9	40.0	42.2
7.Hate speech or slurs against target group(s) outside of Singapore	25.5	30.6	29.4	38.5	31.6	40.7		35.9	44.6	48.1	50.7	35.6	54.4	43.5	31.4	33.3
8. Content endangering public health during a health crisis	37.9	44.0	42.6	52.8	45.2	55.0	64.1		58.9	62.3	64.7	49.6	68.0	57.9	44.9	47.1
9. Misinformation about health risks such as unproven treatments	29.9	35.4	34.1	43.8	36.5	46.0	55.4	41.1		53.5	56.1	40.7	59.7	48.9	36.2	38.3
10. Misuse of inauthentic material e.g. malicious deepfakes without consent	27.0	32.3	31.0	40.4	33.3	42.6	51.9	37.7	46.5		52.6	37.4	56.3	45.5	33.1	35.1
11. False statements about an individual that cause harm or other misleading content	25.0	30.0	28.8	37.9	31.0	40.0	49.3	35.3	43.9	47.4		35.0	53.7	42.9	30.8	32.7
12. Impersonation e.g. identity theft or posing as someone else online	38.3	44.4	42.9	53.2	45.6	55.4	64.4	50.4	59.3	62.6	65.0		68.4	58.3	45.3	47.5
13. Online statements instigating disproportionate harm e.g., starting a cancel campaign	22.3	27.0	25.8	34.4	27.9	36.5	45.6	32.0	40.3	43.7	46.3	31.6		39.2	27.7	29.5
14. Statements affecting reputation i.e., making a statement...that would lower the reputation of a person	30.8	36.4	35.0	44.9	37.5	47.1	56.5	42.1	51.1	54.5	57.1	41.7	60.8		37.2	39.3
15. Misuse of personal information i.e., the unauthorised use of personal information...	42.9	49.1	47.6	57.9	50.3	60.0	68.6	55.1	63.8	66.9	69.2	54.7	72.3	62.8		52.3
16. Doxxing i.e., posting identifiable information about a person with the intent to harass a person.	40.7	46.8	45.4	55.6	48.1	57.8	66.7	52.9	61.7	64.9	67.3	52.5	70.5	60.7	47.7	

4. FACTORS INFLUENCING SEVERITY PERCEPTIONS

This section focuses on the self-reported factors that influenced Singaporeans' severity rankings of online harms. Drawing from insights gathered during the FGDs, we asked survey respondents to reflect on these factors and consider how they shaped their severity perceptions.

From the survey findings, we observed that public perceptions of online harms were primarily shaped through a victim-centric lens. Respondents consistently rated the severity of harm to individuals as the most influential factor, with 73.5 per cent of respondents (M = 4.01, SD = 0.92) indicating the factor to be "very important" or "extremely important" in shaping their perceptions of the severity of online harms. This was closely followed by the vulnerability of potential victims, with 74.3 per cent of respondents (M = 4.00, SD = 0.92) indicating the factor to be "very important" or "extremely important" (see Figure 3). These findings suggest that concern and empathy for affected individuals were central in shaping respondents' severity assessments.

Figure 3: Factors influencing online harms severity



Other influential factors included the presence of existing measures to address the harm, with 71.2 per cent of respondents (M = 3.97, SD = 0.93) indicating the factor to be “very important” or “extremely important”. Similarly, 70.8 per cent of respondents (M = 3.96, SD = 0.93) indicated that the number of people that are likely to be affected by the harm was “very important” or “extremely important” (see Figure 3)

Fewer respondents prioritised the severity of harm to society (M = 3.94, SD = 0.94) or the likelihood of harm happening to someone like the respondent (M = 3.82, SD = 0.95) compared to other factors, with 69.8 per cent and 66.2 per cent of respondents indicating the factors to be “very important” or “extremely important”, respectively (see Figure 3).

These findings were consistent with the FGDs where child sexual exploitation and abuse was ranked as the most severe harm due to the extreme vulnerability of victims and its lifelong consequences.

> "At that age, children have not yet developed their own thoughts or vision of what the world should be. They are particularly susceptible to adopting the views of those who influence them. If they are exploited, especially sexually, they may grow up thinking that such behaviour is acceptable." **— P605, Male, 41 years old, User**

> "The lasting harm [of child sexual exploitation and abuse]... even like when they grow up, their idea of how the world works can be damaged. And at that young age, they are very impressionable, you don't know how entrenched this harm can be." **— P805, Female, 32 years old, Educator**

While FGD participants also raised other factors, such as the scale and duration of impact, countervailing measures and levels of digital literacy, their emphasis was also primarily on the severity of harm to individuals and vulnerability of victims. This reflected a victim-first lens over broader societal consequences.

4.1. Gender Differences in What Influences Severity Perceptions

We also observed statistically significant differences between male and female respondents regarding the factor pertaining to the number of people affected

by the harm[4] — 74 per cent of male respondents rated the factor as "very important" or "extremely important" (M = 4.06, SD = 0.87, p = .01) compared to 67.6 per cent of female respondents (M = 3.86, SD = 0.98). Appendix J shows the results of independent t-tests conducted for all six factors.

One possible explanation could be gender differences in moral reasoning. For example, Friesdorf et al. (2015) found that males have a higher tendency to adopt a utilitarian perspective when faced with a moral dilemma. This means that males' judgements are more likely to be influenced by the number of people affected as opposed to females.

No significant differences were found by age.

5. RESPONSIBILITY OF STAKEHOLDERS IN ONLINE SAFETY

Participants from the FGDs and IDIs expressed a shared view that online safety is a collective responsibility. They highlighted the need for collaborative policymaking that brings together multiple stakeholders (e.g., the government, civil society, industry experts and the wider public) to co-create effective solutions.

> "Legislation is just one institution among many. There are also social institutions and political institutions, which I think are lagging behind.

[4] Independent t-tests were conducted to examine the relationship between gender and importance of the six factors in shaping respondents' perceptions of online harms. Besides the number of people affected by the harm, there were no other statistically significant relationship for gender.

> Legal institutions provide a starting point, but on their own, they are not sufficient." **— P601, Male, 35 years old, User**

> "SHECARES@SCWO... is a one-stop centre for victims of online harm... our purpose is always on supporting victim survivors of online harms in a variety of ways, such as through counselling and emotional support, case management. We offer legal clinics. We also can assist with bringing them down to the police station to just give them the support needed to make a police report... Through the partnerships that has been established on internet platforms, we can also help to provide context in reporting content so that it has a higher... success rate" **— QS05, Female, Online Harm Centre Counsellor**

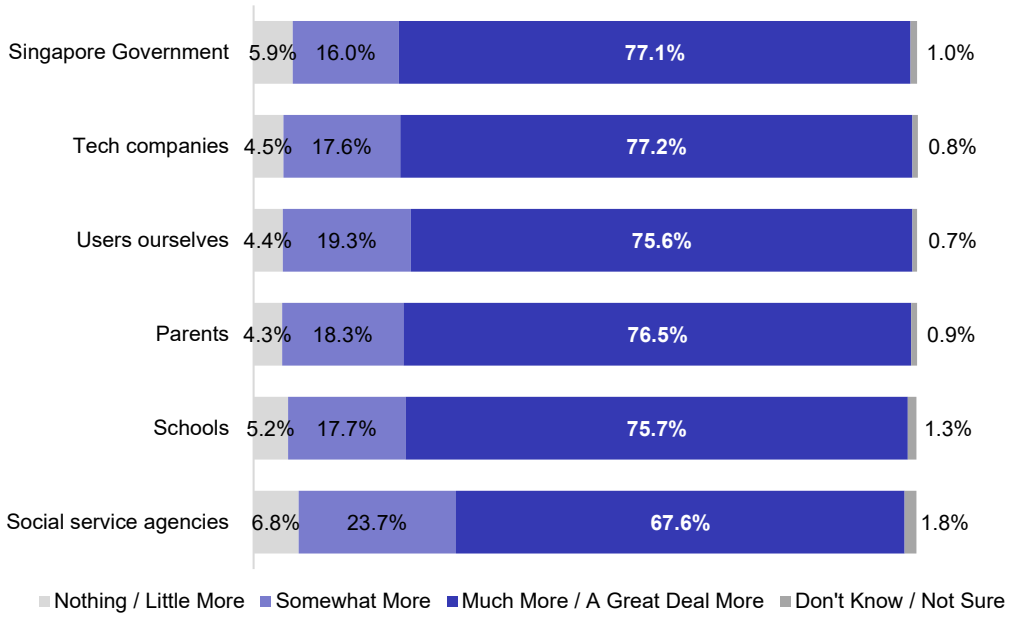
To validate these insights, we asked survey respondents to what extent should each stakeholder — Singapore government, users, tech companies including social media services and app distribution services, parents, schools, and social services agencies and non-governmental organisations — be doing more to improve online safety.

The survey findings revealed strong public consensus that all stakeholder groups have room to do more to improve online safety. Expectations were especially high for systemic actors such as the Singapore government, tech companies including social media and app distributors, and users themselves. The majority of respondents (77.1 per cent, M = 4.14, SD = 0.92) indicated that the Singapore government should do "much more" or "a great deal more".

Similarly, 77.2 per cent of respondents (M = 4.14, SD = 0.88) indicated that tech companies including social media and app distributors should do “much more” or “a great deal more”. Users themselves were also seen as responsible, with 75.6 per cent of respondents (M = 4.14, SD = 0.90) indicating that individuals should do “much more” or “a great deal more”, see Figure 4.

Respondents had the lowest expectations for social service agencies and non-governmental organisations, with 67.6 per cent (M = 3.89, SD = 0.91) indicating that the group should do “much more” or “a great deal more”, see Figure 4.

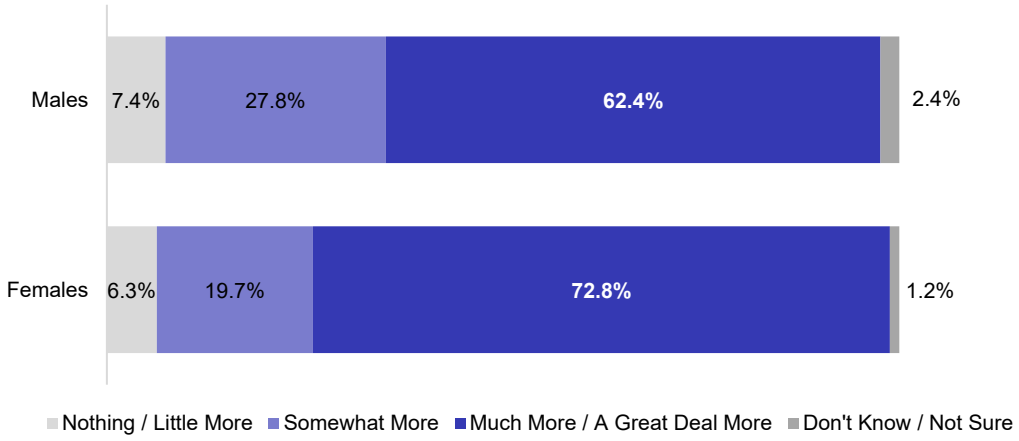
Figure 4: Stakeholders’ responsibility to improve online safety



5.1. Comparing Male and Female Expectations of Stakeholders

We observed statistically significant gender differences regarding the responsibilities by social service agencies and non-governmental organisations (see Appendix K for results of independent t-tests conducted for all stakeholders).[5] Among the female respondents, 72.8 per cent (M = 3.98, SD = 0.87) reported that the stakeholder group should be doing "much more" or "a great deal more" compared to 62.4 per cent of male respondents (M = 3.79, SD = 0.95, p = .01), see Figure 5.

Figure 5: Gender difference regarding expectations of social service agencies



One possible explanation could be gender differences in help-seeking behaviours. Recent research by Goddard and Parker (2025) showed that females are generally more likely to seek mental health support compared to males. This increased likelihood may have contributed to heightened

[5] Independent t-tests were conducted to examine the relationship between gender and expectations of stakeholder's responsibilities to improve online safety. Besides social service agencies, there were no other statistically significant relationship for gender.

awareness about the role that social service agencies and non-governmental organisations play in supporting victims. Thus, shaping female respondents' expectations of these organisations.

This was further supported by our IDIs, where female participants shared detailed experiences with the referral processes and highlighted the need for more victim-centric support.

> "With the polyclinic one, I think it took like, a few weeks or a few months for me to get… an appointment… it took a few months, and by then… I don't really need you that much anymore... this would have been helpful when I was, like, severely depressed... " **— QV08, Female, 30 years old**

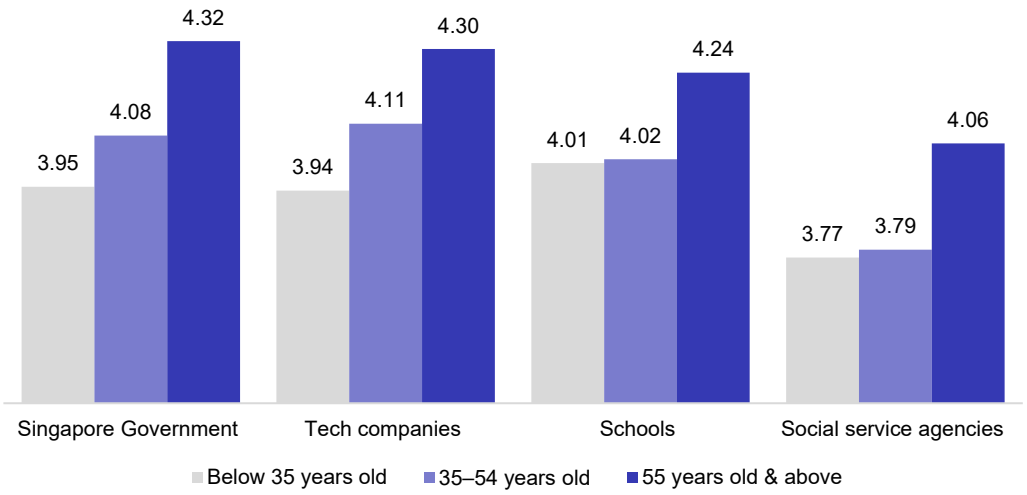
> "I stopped [attending counselling] towards the end… I stopped because my counsellor had to leave the organisation, and they assigned me to another counsellor, which I didn't really… like… so then I decided to not continue." **— QV11, Female, 28 years old**

5.2. Age Differences in Expectations of Stakeholder Responsibility

Age also emerged as a statistically significant factor (Appendix L shows results from the post hoc analyses). In general, respondents aged 55 years old and above had significantly higher expectations of the Singapore government, tech companies, schools, and social service agencies and non-governmental

organisations, compared to the younger groups.[6] See Figure 6 for age differences in mean scores.

Figure 6: Age differences regarding stakeholder's responsibilities based on mean scores



More than eight in 10 older respondents (84.1 per cent, M = 4.32, SD = 0.82) felt that the Singapore government should do "much more" or "a great deal more" as compared to 75.2 per cent middle-aged respondents (M = 4.08, SD = 0.94, $p < .05$) and 69.4 per cent youths (M = 3.95, SD = 0.99, $p < .001$). A similar pattern was observed for tech companies, where 82.6 per cent of older respondents (M = 4.30, SD = 0.79) indicated that tech companies should do "much more" or "a great deal more" as compared to 73 per cent youths (M = 3.94, SD = 0.92, $p < .001$).

[6] An ANOVA was conducted to examine the relationship between age and expectations of stakeholder's responsibilities to improve online safety. Of the six stakeholder groups, only the Singapore government, tech companies, social service agencies and non-governmental organisations, and schools showed significant differences across age groups ($p < .05$). Post hoc analyses were conducted to identify specific group differences.

Older respondents also had higher expectations of schools, with 80.0 per cent of older respondents (M = 4.24, SD = 0.82) indicating that schools should do “much more” or “a great deal more” as compared to 73.1 per cent middle-aged respondents (M = 4.02, SD = 0.96, $p < .05$) and 73.0 per cent youths (M = 4.01, SD = 0.89, $p < .05$).

Finally, 74.3 per cent of older respondents (M = 4.06, SD = 0.84) indicated that social service agencies and non-governmental organisations should do “much more” or “a great deal more” as compared to 62.2 per cent middle-aged respondents (M = 3.79, SD = 0.94, $p < .01$) and 65.3 per cent youths (M = 3.77, SD = 0.96, $p < .01$).

One possible explanation is that older respondents, who tend to have lower digital literacy levels (Infocomm Media Development Authority, 2023a), may feel less confident navigating digital environments. This may lead them to place greater responsibility on these institutions to provide them with protection, support and information.

6. PERCEIVED USEFULNESS OF ONLINE HARMS REMEDIES

To understand Singaporeans’ perceptions about effective measures to tackle online harms, we asked respondents to rate how helpful they perceived various online harms remedies to be on a five-point scale ranging from 1 (“Not helpful”) to 5 (“Extremely helpful”).

Two remedies stood out — nearly eight in 10 respondents (79.3 per cent, M = 4.22, SD = 0.89) deemed it to be “very helpful” or “extremely helpful” when legislation or law holds perpetrators accountable for their actions. Similarly, 77.0 per cent of respondents (M = 4.14, SD = 0.92) deemed it to be “very helpful” or “extremely helpful” when social media services remove harmful content and accounts more quickly. Table 4 shows the mean scores for each remedy and the corresponding proportion of respondents who rated the remedy as “very helpful” or “extremely helpful”.

Table 4: Perceived usefulness of remedies

Remedies	Mean	Standard Deviation	Very / Extremely helpful
When legislation/law holds perpetrators accountable for their actions	4.22	0.89	79.3%
When social media services remove harmful content and accounts more quickly	4.14	0.92	77.0%
When legislation/law can adapt to emerging online harms quickly	4.13	0.85	78.2%
When legislation/law can allow for take down of harmful online content	4.13	0.92	77.4%
When legislation/law empowers victims to seek relief and financial damages for the harm suffered	4.06	0.92	75.0%
When legislation/law is accompanied by public education on the protections conferred by existing laws and measures	3.97	0.88	72.9%
When social media services implement safety by design (i.e., minimisation of risks and harms to users when developing products and services)	3.96	0.89	71.1%
When social media services put in place additional safety measures to protect children (e.g., age verification measures)	3.98	0.94	72.3%
When social media services improve user reporting tools and processes	3.93	0.92	70.6%
When legislation/law ensures that individuals are aware of their responsibilities online, and penalises them for not fulfilling their responsibilities	3.94	0.95	69.7%

When public education efforts and campaigns are relatable and incorporate real-life stories	3.90	0.90	69.6%
When public education efforts and campaigns tell me what I should do by providing step-by-step processes on actions to take when I encounter online harms, or if I think I have encountered online harms	3.90	0.93	69.0%
When public education efforts and campaigns tailor content to my age group and life stage	3.90	0.95	67.7%
When legislation/law offers ways to help offenders reform and improve	3.83	0.97	67.6%
When social media services publish transparency accountability reports	3.85	0.98	65.7%
When public education efforts and campaigns deliver messages through creative and engaging formats	3.79	0.96	62.9%

Legislative remedies	Action by social media services	Public education

Other legal remedies were also deemed very useful. More than 70 per cent of respondents considered it to be “very helpful” or “extremely helpful” when legislation or law can adapt to emerging online harms quickly, remove harmful content, empower victims to seek relief and financial damages, and is accompanied with public education on the existing legal protections.

Insights from the IDIs reinforced these findings regarding the top three most useful remedies. Participants emphasised the need for legislations to ensure perpetrators are held accountable and reflect the evolving nature of online harms. Additionally, for victims, the swift removal of harmful content was a clear priority, underscoring the critical role of timely and effective content moderation. This was evident in the case of QV11, whose intimate videos were uploaded online without her consent by an ex-boyfriend:

> "After... I found out. I did a few things. One was to try and... download the video using VPN... so that it doesn't get as many views or traction. Also check to report... on that website. So I tried to do that for that video... there was not only one video, several videos... I tried to report it." — **QV11, Female, 28 years old**

FGD findings supported the survey results whereby participants highlighted the need for legislation to be both ex-post (reactive) and ex-ante (preventive). Many noted that harm often occurs before laws take effect, reducing public confidence in legislation as a meaningful form of protection. This is exacerbated by the slow pace of legislative reform, which frequently lags behind the rapid evolution of digital platforms and behaviours. Taken together, the survey findings offer strong support for the remedies covered in the forthcoming Online Safety (Relief and Accountability) bill.[7] [8]

Comparatively, non-legislative interventions were perceived as somewhat less helpful. For example, while 65.7 per cent of respondents (M = 3.85, SD = 0.98) rated increased transparency and accountability from social media services as "very helpful" or "extremely helpful," this was lower than any of the legal remedies. Although transparency was frequently raised in FGDs and IDIs,

[7] The Online Safety (Relief and Accountability) bill aims to provide stronger assurance to victims and raise accountability standards. The proposed measures under this bill include: (1) a new Online Safety Commission to receive and swiftly act on victim's complaints, (2) new statutory torts for civil claims against perpetrators, and (3) enhanced user information disclosure to aid investigations (Ministry of Law, 2024).

[8] Public consultation outcomes: More than 90 per cent of respondents strongly supported the creation of a dedicated agency to address online harms; more than 95 per cent of respondents were in favour of granting victims the right to pursue legal action against perpetrators; about 80 per cent of respondents agreed that it would be useful to disclose a perpetrator's user information to the victim to improve accountability and deter online harms (REACH, 2025).

survey respondents perceived it as less helpful. One explanation is the interpretation of what transparency entails. For many FGD and IDI participants, it meant clearer and more consistent communication from platforms about how content moderation decisions are made — especially why certain content is removed or retained.

> "Sometimes we wear very revealing clothes... every time IG likes to flag me or say that my content is actually explicit… but you allow it on other people's posts (unconsented posting of the victim) and then when we report, nothing happens, [it's] like the double standards, right." **— QV12, Male, 31 years old**

> "We've also had clients who had their intimate photos on Instagram. When they reported it on the app, it wasn't removed, which was very strange, because it's literally nudity." **— QS05, Female, Online Harm Centre Counsellor**

Interestingly, 67.6 per cent of respondents (M = 3.83, SD = 0.97) deemed it to be "very helpful" or "extremely helpful" when legislation or law offers ways to help offenders reform and improve. While FGDs and IDIs highlighted the value of helping offenders understand the impact of their actions and encouraging behavioural change, survey results suggested such remedies to be less helpful than punitive or victim-focused measures. This could be due to low public understanding of offender-centric interventions and the restorative justice approach (Restorative Justice Circle, 2016). Some people also think it is not

suitable for certain harms (The Asia Foundation, 2023) and people could be considering the impact of online harms from the victim's lens and hence not see the broader significance and utility of rehabilitation (Wenzel et al., 2008).

6.1. Age Differences in Ratings of Online Harms Remedies

Our analysis found that age is a predictor of perceived usefulness of online harms remedies. Respondents aged 55 years old and above consistently rated these remedies as more helpful than younger age groups. Across most remedies, older adults gave mean ratings above 4.00, with more than 70 per cent of this age group considering the remedies "very helpful" or "extremely helpful" [9] Appendices M to O show statistically significant results from post hoc analyses).

Statistically significant results were observed in responses to specific remedies. For example, 84.5 per cent older of respondents (M = 4.36, SD = 0.76) deemed it to be "very helpful" or "extremely helpful" when legislation or law holds perpetrators accountable, compared to 76.1 per cent respondents aged 35–54 years old (M = 4.14, SD = 0.93, p = .03) and 76.0 per cent respondents below 35 years old (M = 4.12, SD = 0.98, p = .02) (see Appendix O). This could be due to the higher confidence older adults have in legal institutions as compared

[9] An ANOVA was conducted to examine the relationship between age and perceived usefulness of various online harms remedies. Out of the 16 remedies that respondents rated on, no significant results (p >.05) were found for these remedies: (1) public education efforts and campaigns delivered messages through creative and engaging formats, (2) public education efforts and campaigns are relatable and incorporate real-life stories, (3) when legislation/law can adapt to emerging online harms quickly, (4) when legislation/law offers ways to help offenders reform and improve, and (5) when legislation/law ensures that individuals are aware of their responsibilities online, and penalises them for not fulfilling their responsibilities. Post hoc analyses were conducted for the remaining remedies to identify specific group differences.

to younger age groups (Hamm et al., 2016). Similarly, 81.5 per cent of older respondents (M = 4.22, SD = 0.84) deemed it to be “very helpful” or “extremely helpful” when legislation or law empowers victims to seek relief and financial damages for the harm suffered, compared to 69.3 per cent of middle-aged respondents (M = 3.97, SD = 0.92, p = .01) and 73.0 per cent of youths (M = 3.97, SD = 0.98, p = .02).

This age difference was also evident in responses to public education efforts. 78.5 per cent of older respondents (M = 4.10, SD = 0.82) deemed it to be “very helpful” or “extremely helpful” when public education efforts and campaigns provide step-by-step processes on actions to take when they encounter online harms, or if they think they have encountered online harms, compared to 63.9 per cent of middle-aged respondents (M = 3.81, SD = 1.00, p = .003) and 62.3 per cent of youths (M = 3.74, SD = 0.93, p < .001) (see Appendix N).

Older respondents also perceived certain remedies by social media services to be more helpful compared to other age groups. For example, 79.5 per cent of older respondents (M = 4.13, SD = 0.83) deemed it to be “very helpful” or “extremely helpful” when social media services implement safety by design, compared to 69.3 per cent of middle-aged respondents (M = 3.91, SD = 0.91, p < .05) and 61.7 per cent of youths (M = 3.76, SD = 0.90, p < .001) (see Appendix M). Additionally, 82.1 per cent of older respondents (M = 4.19, SD = 0.86) deemed it to be “very helpful” or “extremely helpful” when social media services put in place additional safety measures to protect children compared

to 68.1 per cent of middle-aged respondents (M = 3.88, SD = 0.96, p = .001) and 64.0 per cent of youths (M = 3.82, SD = 0.99, p < .001).

The consistent endorsement by older respondents may partly reflect their lower digital literacy levels and lower self-efficacy. Limited exposure to online platforms may have led to less critical engagement with specific interventions, resulting in a general tendency to view all proposed solutions as helpful.

Perceived usefulness of remedies did not differ significantly by gender.

7. EXPERIENCES AND FUTURE OUTLOOK

We also examined Singaporeans' experiences of online harms and their perceptions of the future of the digital landscape.

7.1. Online Harms Experience

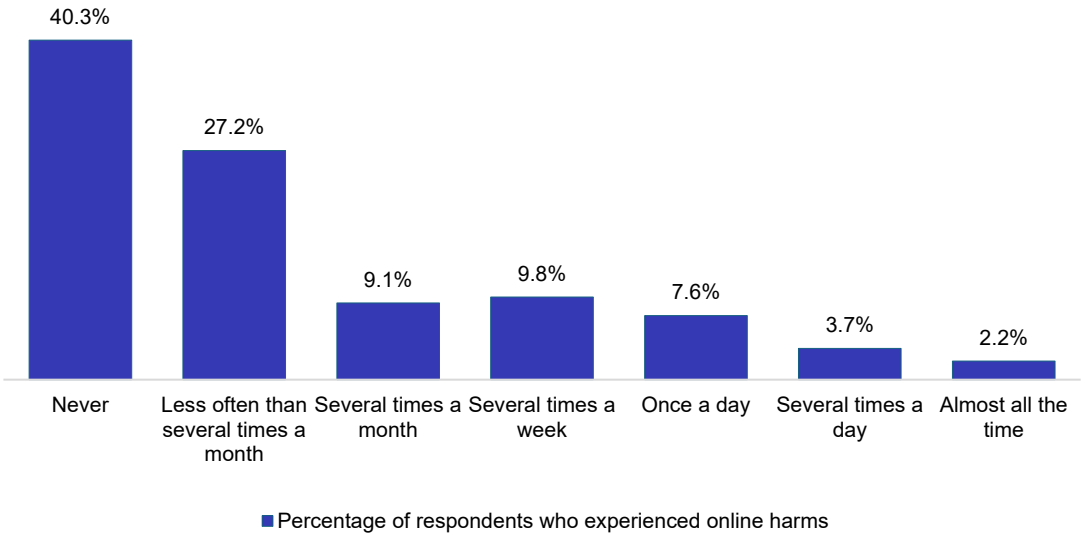
To measure respondents' experiences of online harms, we asked them to indicate their encounter with online harms over the past 12 months using a seven-point frequency scale ranging from 1 ("Almost all the time") to 7 ("Never").[10]

The majority of respondents (60 per cent) reported that they had experienced online harms at varying frequencies (see Figure 7). These findings are

[10] The data was reverse coded for analyses purposes.

consistent with national data on online harms in Singapore, where prevalence rates ranged from 58 per cent to 66 per cent between 2023 and 2024 (Ministry of Digital Development and Information, 2024; Ministry of Digital Development and Information, 2023; SG Her Empowerment, 2023).[11]

Figure 7: Online harm experience



7.1.1. How gender affects exposure to online harms

Our survey findings point to gender as a predictor, with females being slightly less likely to experience online harms as compared to males. Specifically, the mean score for frequency of encountering online harms was higher for male respondents (M = 2.54) than for female respondents (M = 2.21). Appendix P shows results from an independent t-test. This aligned with findings from SG Her Empowerment which found that more males than females reported personally experiencing online harms (SG Her Empowerment, 2023).

[11] Prevalence of encountering online harms: 2023 study by SG Her Empowerment (SHE): 58 per cent; 2023 MDDI study: 65 per cent; 2024 MDDI study: 66 per cent.

7.1.2. How age affects exposure to online harms

Age was also as a predictor of online harm frequency, with older respondents reporting less frequent encounters compared to middle-aged and youth participants. The mean score for frequency of encountering online harms was the lowest for respondents aged 55 years old and above (M = 1.98), compared to those for respondents aged 35 years old and below (M = 2.63) and those aged 35–54 years old (M = 2.59). Appendix Q shows the results of post hoc analysis conducted to examine age differences.

These findings aligned with those from the FGDs where older participants felt less susceptible to online harms, often dismissing others' opinions online or ignoring cyberbullying attempts more easily than younger individuals.

> "The primary social lives of especially younger people are now online. My social life is not online. I think that makes a very big difference. The kind of bullying I faced when I was younger was face-to-face… I don't have an open record of it online [sic]. That's not the case for the younger generation." **— P601, Male, 35 years old, User**

> "Grooming of young kids, cyberbullying, people leaving very mean comments, judging someone by their appearance, telling them how they ought to be, cussing, asking them to die, stuff like that… younger kids

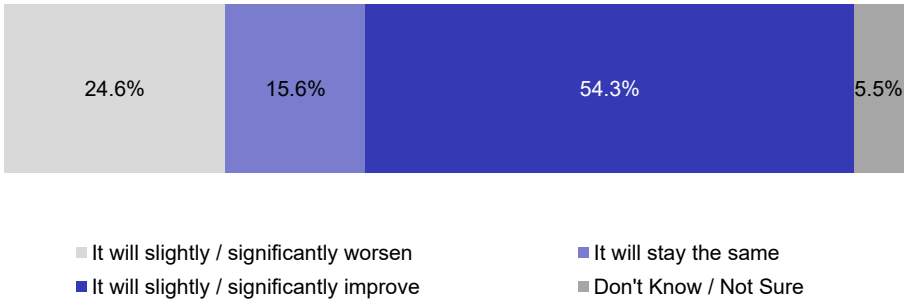
who are very impressionable and unable to think as clearly as adults might take these things too hard." **— P105, Female, 23 years old, User**

Additionally, the findings were consistent with broader prevalence patterns of online harms across age groups, with prevalence rates decreasing as age increases (Enock et al., 2023). One possible explanation could be the amount of time spent online. In the study by Enock et al. (2023), younger individuals were found to spend significantly more time online compared to older adults, which contributed to their higher exposure to online harms.

7.2. Perceptions of the Future of Online Safety

We assessed respondents' perceptions of the future of online safety by asking them to rate on a five-point scale, ranging from 1 ("It will significantly worsen") to 5 ("It will significantly improve"). Overall, more than half of respondents (54.3 per cent) were optimistic, believing online safety would "slightly or significantly improve" (see Figure 8). About one in four (24.6 per cent) were pessimistic.

Figure 8: Perceptions of online safety



A possible explanation for respondents' optimism could be due to priming effects, a phenomenon in which exposure to one stimulus influences how a person responds to a subsequent, related stimulus (American Psychological Association, 2018). Since respondents were asked about the effectiveness of remedies prior to this question—most of which were perceived as helpful—this likely shaped their perceptions of online safety in a more positive direction.

8. RECOMMENDATIONS

Our study found that scams and cybercrimes, child sexual exploitation and abuse, violent or terrorism-related content, and content supporting vice and organised crime are top concerns among Singaporeans. These are followed by non-consensual sexual content, promotion of dangerous behaviours, and targeted harassment. These perceptions reinforce how victim-centric concerns — such as the extent of harm and the vulnerability of victims — shape public perceptions of harm severity.

The study clearly points out that the responsibility for improving online safety does not rest solely with the Singapore government or tech companies. Instead, it is seen as a shared responsibility among multiple stakeholders.

Drawing from insights across all study phases, we propose the following recommendations to address online harms. These have been organised into three parts: (1) what is working and should be reinforced, (2) what remains

lacking and needs to be addressed, and (3) what we must anticipate as new and emerging challenges.

8.1. What is Working and Should be Reinforced

8.1.1. Enhance victim support pathways

Singapore has introduced key legislations in recent years to improve online safety, including the Online Safety (Miscellaneous Amendments) Act 2022 and the Online Criminal Harms Act 2023. These laws require social media platforms to take active steps to limit the spread of harmful content, such as disabling access, restricting harmful accounts, and putting in place victim support mechanisms. Under the Code of Practice for Online Safety – Social Media Services, designated social media services must also implement system-wide measures to minimise users' exposure to harmful content and provide users accessible mechanisms to report it (Infocomm Media Development Authority, n.d.).

However, findings from our IDIs revealed that victims continue to face barriers when seeking help. Many shared that they were unclear about what would happen after they report and were hesitant to come forward due to fears of burdening others, not being believed, or facing reputational harm and unwanted exposure.

> "Initially, I wanted to stop [the perpetrator/online harm] myself because I don't want to trouble people. When you trouble people, then you still must tell them all the background info. Then what if they don't believe me?" **— QV09, Female, 31 years old**

> "I was terrified that if I reported it to the police, it would somehow leak to… the press... I really, really didn't want my name out there." **— QV08, Female, 30 years old**

A public report released in February 2025 (Infocomm Media Development Authority, 2025) reinforced these concerns, pointing to delays in platform response, as well as gaps in public understanding of about reporting procedures and organisational thresholds for action. This lack of clarity and confidence can deter victims from coming forward to seek help and discourage bystanders from intervening.

The establishment of the Online Safety Commission in 2026 presents an opportunity to close these gaps. Practical models such as the Australian eSafety Hotline (eSafety Commissioner, 2015) and the UK's Revenge Porn Helpline (Revenge Porn Helpline, n.d.) offer workable examples of how confidential guidance and rapid escalation protocols can be implemented.

Based on our findings, the Commission should prioritise two key areas in the short to medium term: (1) demystify the reporting process and clarify institutional roles and responsibilities of different stakeholders; (2) develop

accessible, multilingual resources that outline clear steps for filing complaints, expected timelines and available support services.

8.1.2. Adapt legislation to address emerging online harms

Technology is evolving rapidly, and legislation addressing online harms must keep pace. While Singapore's current laws provide a strong foundation, emerging technologies like artificial intelligence have introduced new challenges, particularly around non-consensual synthetic intimate imagery, such as deepfake pornography. In fact, survey respondents explicitly identified deepfakes as a severe harm in the open-ended section of our survey. Although existing laws such as the Penal Code 1871 (amended 2012), the Films Act 1981 (amended 2019) and the Protection from Harassment Act 2014 can be applied to take offenders to task, they do not specifically address the unique nature and scale of AI-generated content.

Our findings suggest a need for legal reforms that explicitly criminalise deepfake pornography, with provisions that do not require proof of intent or awareness of harm. Such reforms should also introduce stricter offender liability and provide clear legal remedies, including immediate takedown orders and civil damages for victims.

Countries like Australia (Middleton, 2024) and South Korea (Lee, 2024) are already taking steps in this direction, introducing targeted legislation and enhancing cross-border enforcement. Singapore can similarly update its legal framework by introducing specific laws targeting deepfake pornography and

establishing clear, accessible and timely processes for content removal and victim support. Strengthening international collaboration will also be crucial to effectively address the global nature of these harms.

8.1.3. Tailor and target digital literacy initiatives better

Our study points to varying experiences and needs among different demographic groups when it comes to online safety. For example, females and youths are more concerned about the misuse of inauthentic material and cyberbullying, and males and older individuals with hate speech and impersonation. Their anxieties and worries often stem either from their personal experiences or those of their peers, underscoring the need for tailored public education.

Singapore has already made significant strides in promoting digital literacy through initiatives like IMDA's Digital for Life, the Ministry of Education's Cyber Wellness curriculum in schools, and programmes by agencies like National Library Board, Singapore Police Force and Cyber Security Agency. These programmes generally focus on explaining different types of harm, their consequences and how individuals can respond.

Our study findings highlight the need for harm-specific education to be tailored by age and gender. For instance, Internet Matters offers material on pornography, and self-esteem and body image for children as young as six to 10 years old (Internet Matters, n.d.). Similarly, Singapore can develop tailored modules and materials for schools, workplaces and community centres that

address the most relevant harms for each group, thereby increasing their resonance and effectiveness. A working paper by the first and second authors discusses a heuristic for enhancing public education on online safety (Chew et al., 2025).

Parents, educators and social service professionals from the FGDs also expressed concerns over the technical literacy gap between young people and the adults responsible for guiding them. This gap leaves young people more vulnerable to online harms.

> "Nowadays, [children] use incognito mode more often since parents can see which websites [their children] visit and how often, through parental controls. So, [children] simply use incognito mode to avoid detection." **— P704, Female, 44 years old, Parent**

> "Parents can set parental controls on phones and all kinds of safety measures. I don't know how, but sometimes [children] are able to bypass it [safety measures]." **— P807, Male, 37 years old, Educator**

Existing initiatives must therefore be ramped up to include actionable tool-based tips for parents, educators and supporters. Beyond government efforts, industry partners and social service agencies should collaborate to offer more comprehensive digital literacy training for parents and educators. This training should go beyond prevention to include practical guidance on how to respond and provide support after an incident occurs.

8.1.4. Strengthen collaborations within the ecosystem

A key feature of Singapore's approach to managing digital risks is its emphasis on government-industry collaboration. This was evident in the national response to online falsehoods between 2017 and 2019 where the government consulted a wide range of stakeholders including industry players like social media services to develop a comprehensive strategy that included platform-level interventions for debunking falsehoods.

A similar collaborative approach is now being applied to online harms. One example is the requirement for designated social media services to submit annual online safety reports, outlining the measures they have taken to mitigate harm and improve user safety (Infocomm Media Development Authority, n.d.).

Moving forward, institutionalising joint working groups involving policymakers, social media services and non-governmental organisations would ensure that emerging digital threats are regularly reviewed and that response protocols are updated in a timely and coordinated manner. This will enable Singapore to remain adaptive and responsive to fast-evolving online environments.

Such multi-stakeholder partnership is especially important as Singapore's legislations and strategies are closely observed and sometimes emulated by its regional neighbours. By encouraging stronger collaborations within the ecosystem, Singapore can potentially influence better practices across the

region, which in turn strengthens its own defences against transboundary harms.

8.2. Address What's Lacking

8.2.1. Address ambiguities to improve reporting and enforcement, and correct misconceptions

Our study highlights persistent challenges in providing consistent and accessible support for victims of online harms. Findings from our IDIs, alongside a recent report by SG Her Empowerment (2025) on the experiences of survivors, highlight structural gaps in seeking redress for online harms such as inconsistent court procedures, unclear law enforcement protocols and varied interpretations of harm among social service agencies.

> "On the day of the mediation, it was actually like the judge himself, so he actually granted the protection order on the spot. So there was actually no mediation. I was a bit confused, because I was asking him… after I don't want to settle this, can I proceed with the PO (Protection Order)? And he was like… I can just give you the PO now." **— QV03, Female, 30 years old**

> "You have different judges with different takes on things. Some judges will say… naked photos on internet. Serious. Expedited order. Some judges will say, is anybody going to die? No, right? No expedited order." **— QS06, Male, Lawyer**

While most of these accounts were anecdotal, they point to a broader pattern of victim uncertainty that can undermine their confidence in seeking redress. Overly broad or vague definitions of online harm may hinder effective intervention and cause confusion among both victims and supporters. This is evident in cases like doxxing, where the severity and context of harm vary widely. FGDs also surfaced confusion over harms like impersonation and sexual content. For instance, “impersonation” ranged in perception from harmless pranks to serious identity theft, while “sexual content” spanned consensual sharing to exploitative or non-consensual distribution.

The effectiveness of existing legislation is further compromised by low public awareness and misperceptions. Many individuals were unaware of their rights or lacked confidence in the availability or effectiveness of legal remedies. Such misconceptions and confusion can compound emotional distress and deter them from seeking help. Ambiguity surrounding institutional thresholds like what constitutes actionable harm can also hamper the ability of social service agencies to render support.

To address these issues, policymakers and platforms should publish clear, regularly updated guidelines that outline reporting procedures, assessment criteria and possible outcomes. The Office of the eSafety Commissioner in Australia offers a useful model, using flowcharts and case studies to help users understand when and how to seek help (eSafety Commissioner, 2023). Clearer guidelines will also enable social service agencies to align their practices and

assessment criteria, promoting a more unified and consistent approach to case management.

Equally important is effective public communication. Legislative provisions must be explained through accessible channels, using plain language and relatable examples. A useful precedent is the layperson-friendly fact sheets (Intellectual Property Office of Singapore, 2025) issued alongside the Copyright Act amendments in 2021. The government and the forthcoming Online Safety Commission could partner with social service agencies to co-develop simplified guides that show what the law covers, how individuals can protect themselves and steps involved in seeking redress.

Outreach efforts should also leverage social media services and community networks to reach different demographic groups, particularly the less digitally literate populations. By addressing misconceptions, clarifying procedures, and highlighting success outcomes, these initiatives can empower more victims to seek help and enhance the deterrent effect of the law.

8.2.2. Improve platform responsiveness and transparency

Distrust in technology platforms, often stemming from perceived inaction and a lack of transparency, can result in user apathy and reluctance to use available reporting tools. As platforms are typically the first point of contact for victims of online harms, it is imperative that they improve both the speed and transparency of their responses.

To rebuild trust, the industry should publish regular reports detailing how cases are handled, including data on response times, outcomes and types of interventions (even though respondents in our study viewed this measure as less helpful compared to other legislative and platform-based actions). The government has already provided feedback on initial reports from the six designated social media services, highlighting areas for improvement. Policymakers should build on this by mandating minimum standards for platform responsiveness and transparency going forward.

In addition to transparency, platforms should be encouraged to adopt safety-by-design practices that reduce risk exposure at the interface level. Design choices such as content demotion, friction prompts, and youth-specific safety settings can play a preventive role by reducing the spread of harmful content and encouraging more mindful user behaviour. Embedding these safety mechanisms upstream complements downstream measures like user reporting and moderation, creating a more resilient digital environment.

There are also opportunities for co-developing joint training programmes by government agencies, industry and social service agencies. These programmes should target frontline responders, educators, community leaders and platform moderators. Such partnerships will foster a shared understanding of local harm contexts, legal thresholds for intervention and the operational protocols needed to close current response gaps.

8.2.3. Create a unified national platform for sharing best practices

The current landscape of efforts to tackle online harms is fragmented, with various stakeholders — government agencies, industry and the social service sector — often operating in parallel and sometimes duplicating efforts. This lack of coordination hampers efficient knowledge transfer and prevents the scaling of successful interventions.

A unified national platform, similar to the European Union's Better Internet for Kids portal (Better Internet for Kids, n.d.) could provide a centralised space for the systematic sharing of data and best practices. It would also facilitate collaboration among stakeholders with different expertise and resources, strengthening collective responses to online harms.

Policymakers should spearhead the development of such a platform, ensuring it is accessible, regularly updated and jointly supported by all key stakeholders. The platform could serve as a one-stop resource for victims, educators and service providers and host materials such as:

- Case studies and anonymised incident summaries
- Toolkits for prevention and response
- Legislative updates
- Educational resources tailored for different user groups

This initiative should be structured as a collective endeavour. For example, industry partners could contribute anonymised data and platform-specific insights; social service agencies could offer practical toolkits and provide feedback on clients' experiences and regulators could update the legal and procedural content in line with policy changes.

By overcoming silos and fostering ongoing exchange, a national platform would not only improve coordination and capacity-building but also encourage the development of more innovative and responsive solutions, ultimately strengthening the overall ecosystem for tackling online harms.

8.2.4. Build a culture of shared responsibility

Addressing online harms starts with building a culture of shared responsibility—where individuals, family, peers in schools and workplaces all play an active role in prevention and intervention. Our FGDs revealed how behaviours like catfishing are often normalised, especially among young females who described them as a form of harmless fun.

> "When I was younger, at the age of 15 years old, I had a more playful online personality. I probably said a lot of things that I wouldn't say in real life. I found a lot of joy in catfishing because I was playing with my friends. In online games, I was urged by my friends to have a little fun

> with a guy who was desperate [sic]. So I put on a different personality, the type that we think men want." **— P106, Female, 23 years old, User**
>
> "At that age, we were all trying to find ourselves [sic]. There is a domino effect as every person of that age will do similar things. When I was 15 years old, my friend encouraged me to use Tinder so that we could catfish people. We would flirt with old men and it was quite funny." **— P103, Female, 20 years old, User**

Educational campaigns must directly challenge such attitudes, especially those that trivialise behaviours like catfishing and clout chasing. Personal stories and victim testimonials can serve as powerful tools to humanise the impact of harm and counter the distancing effects of "othering"[12]. In the UK, the "Stop, Speak, Support" campaign (Anti-bullying Alliance, n.d.) encourages children and teenagers aged 11 to 16 years to be responsible digital citizens and to help stop the spread of cyberbullying.

Beyond awareness campaigns, it is important to involve youth directly in shaping interventions. Co-designing initiatives with youth — through advisory panels, peer-led programmes, or school-based workshops — can ensure interventions are not only relevant but also empowering. Youth can also play meaningful roles as digital ambassadors, helping to shift peer norms and promote responsible behaviour from within their own networks.

[12] Othering is the belief that online harms are problems faced by others.

Workplaces offer an important avenue for outreach. Young adults and middle-aged individuals who have aged out of school-based digital literacy initiatives remain at risk. They are also likely to be peers or parents to victims. Initiatives like the workshops conducted by the Institute of Strategic Dialogue in the UK which facilitate structured conversations about recognising and responding to hate speech and extremism in workplace settings, offer a potential model for adaptation in Singapore.

Ultimately, fostering a culture of collective responsibility will not only deter perpetrators but also ensure that victims receive timely support from all segments of society.

8.3. Address What's Emerging

8.3.1. Offender-centric approaches — Rehabilitation and restorative justice

Emerging research reveals that some perpetrators, especially youths, may act out of ignorance rather than malice because they lack awareness of the impact of their behaviour. FGDs indicate that young girls, for example, sometimes engage in acts like catfishing, perceiving them as harmless pranks. These insights point to the potential of offender-centric approaches that emphasise rehabilitation and restorative justice, offering alternatives to punitive-only models.

Policymakers should collaborate with educators, psychologists, and legal experts to design diversion programmes that focus on education, empathy-building and reconciliation. For instance, the Pathways to Accountability, Change and Transformation (Pact) programme (Ang, 2025) conducts pre-sentencing intervention aimed at breaking cycles of harm and supporting long-term reintegration. Since its launch in January 2024, Pact has supported 97 offenders out of whom 41 committed sexual crimes like upskirt filming.

Singapore's juvenile justice system which already prioritises rehabilitation can be further adapted to address digital offences. In school settings, restorative practices in response to cyberbullying and online harassment have shown promising results, including lower suspension rates, better academic outcomes and an improved school environment. These programmes (Zizzola, n.d.) centre on values such as empathy, accountability, respect, and acceptance.

International models such as New Zealand's restorative justice framework (Ministry of Justice, 2022) offer valuable guidance. Voluntary facilitated meetings bring together victims, offenders, support people and other approved people (e.g., community representatives and interpreters) to address harm, promote accountability, and agree on meaningful redress. Offenders are given the opportunities to acknowledge their actions, apologise and take steps to put right the harm they have caused. This process supports both victim closure and offender rehabilitation and reduces recidivism.

While such initiatives may not be suitable for all types of online harms, they offer an important complement to traditional penalties. They must be carefully managed to ensure victim needs are prioritised and that safeguards are in place. Social service agencies and community groups can support implementation through counselling and mediation, while the tech industry can develop educational modules for users who breach community standards.

By adopting offender-centric approaches, Singapore can better address root causes of online harm and promote longer-term behavioural change.

8.3.2. Cross-border online harms

The borderless nature of the internet presents persistent challenges, especially when perpetrators operate from jurisdictions with limited legal infrastructure or weak enforcement mechanisms. To address this, Singapore should prioritise bilateral and multilateral agreements that facilitate information sharing, investigation, and prosecution of cross-border digital crimes. The arrests of 400 over people for online child sexual exploitation activities in February and March 2025 by law enforcers from Singapore, Hong Kong, Japan, South Korea, Malaysia and Thailand is an example of cross-border information sharing and coordination (Hamzah, 2025).

Singapore can build on existing platforms like the ASEAN Ministerial Conference on Cybersecurity to expand regional cooperation to tackle online harms (Minárik, n.d.). At the international level, the Council of Europe's

Budapest Convention on Cybercrime offers a model for cross-border cooperation, including standardising protocols for evidence collection and mutual legal assistance (Council of Europe, n.d.).

Initial efforts to strengthen cross-border collaboration can focus on clear-cut domains where the problems see common definition, like scams, child sexual abuse content, and non-consensual sexual imagery. These areas provide a practical starting point for enhancing regional alignment and building operational capacity.

Stronger regional and international coordination will help close protection gaps for victims and exert collective pressure on social media services to respond to cross-border data requests, enable content takedowns, and implement geo-blocking where necessary.

9. CONCLUSION

As digital technologies continue to evolve, so too will the forms, intensity and reach of online harms. To keep pace, policymakers should ensure that legislation remains robust and adaptive to reflect emerging threats. At the same time, technology companies must be held accountable for timely responses and transparent moderation practices.

Public awareness campaigns should clearly outline the specific responsibilities of each stakeholder and use real-life examples to illustrate how coordinated

action can prevent and address harm. The industry must strengthen its existing systems for user feedback, appeals and reporting, and provide regular transparency reports that show how they are addressing localised risks and concerns.

The social service sector will need to continually update risk assessment tools and protocols, and share insights with policymakers and industry to close gaps in intervention and support. Finally, users themselves must be equipped with the knowledge and tools to protect themselves and support others.

Together, these efforts will foster a digital environment built on shared responsibility, where safety is not delegated to a single actor, but sustained through collective action.

REFERENCES

American Psychological Association. (2018, April 19). *Priming*. In APA Dictionary of Psychology. Retrieved May 27, 2025, from https://dictionary.apa.org/priming

Amnesty International. (2017, November 20). *Amnesty reveals alarming impact of online abuse against women*. https://www.amnesty.org/en/latest/press-release/2017/11/amnesty-reveals-alarming-impact-of-online-abuse-against-women/

Ang, S. (2025, May 19). Pre-sentencing therapy helps offenders with underlying issues to break cycle of reoffending. *The Straits Times.* https://www.straitstimes.com/singapore/pre-sentencing-therapy-helps-offenders-with-underlying-issues

Anti-bullying Alliance. (n.d.). Stop Speak Support: Challenging online bullying. https://anti-bullyingalliance.org.uk/tools-information/all-about-bullying/online-bullying/stop-speak-support-challenging-online-bullying

Baltezarevic, R., & Baltezarevic, I. (2024). *Social media impersonation as a cybersecurity threat.* International Topkapi Congress-IV, Istanbul, Türkiye. https://www.researchgate.net/publication/384657784_SOCIAL_MEDIA_IMPERSONATION_AS_A_CYBERSECURITY_THREAT

Bestvater, S., Gelles-Watnick, R., Odabaş, M., Anderson, M., & Smith, A. (2023, June 29). *Americans' views of and experiences with activism on social media*. Pew Research Center. https://www.pewresearch.org/internet/2023/06/29/americans-views-of-and-experiences-with-activism-on-social-media/

Better Internet for Kids. (n.d.). *About*. https://better-internet-for-kids.europa.eu/en/about

Chew, H. E., Lim, S. S., & Soon, C. (2022, November 15). Online safety rules must be matched by public education. *The Straits Times.* https://www.straitstimes.com/tech/tech-news/online-safety-rules-must-be-matched-by-public-education

Chew, H. E., Soon, C. & Chia, N. (2025, July 14). *Borrowing Berlo: Enhancing public understanding of Singapore's online harms laws through the SMCR model of communication* (IPS Working Paper No. 63). Institute of Policy Studies, Lee Kuan Yew School of Public Policy, National University of Singapore. Retrieved from https://lkyspp.nus.edu.sg/ips/news/details/ips-working-papers-no.-63---borrowing-berlo-enhancing-public-understanding-of-singapore's-online-harm-laws

Council of Europe. (n.d.). *The Convention on Cybercrime (Budapest Convention, ETS No. 185) and its Protocols.* https://www.coe.int/en/web/cybercrime/the-budapest-convention

eSafety Commissioner. (2023). *Summary table of what you can report and how.* https://www.esafety.gov.au/report-online-harm/summary-table-of-what-you-can-report-and-how

eSafety Commissioner. (2015, September 11). *Protecting children online, one image at a time* [Press release]. https://www.esafety.gov.au/newsroom/media-releases/protecting-children-online-one-image-time

European Training Foundation. (2020). *Mapping Covid 19*. https://www.etf.europa.eu/sites/default/files/2020-07/etf_covid_mapping_v06_1.pdf

Enock, F., Johansson, P., Bright, J., & Margetts, H. (2023). *Tracking experiences of online harms and attitudes towards online safety interventions.* The Alan Turing Institute. tracking_experiences_of_online_harms_and_attitudes_report_final.pdf

Films Act 1981 (2020 revised edition). https://sso.agc.gov.sg/Act/FA1981

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in responses to moral dilemmas: A process dissociation analysis. *Personality & social psychology bulletin, 41*(5), 696–713. https://doi.org/10.1177/0146167215575731

Goddard, I., & Parker, K. (2025, January 16). *Where men and women turn for emotional support and social connection.* Pew Research Center. https://www.pewresearch.org/2025/01/16/where-men-and-women-turn-for-emotional-support-and-social-connection/

Goh, Y. H. (2024, April 24). “Black hole” of alluring content: Chinese social media apps like Xiaohongshu pull in S’pore users. *The Straits Times*. https://www.straitstimes.com/asia/east-asia/black-hole-of-alluring-content-chinese-social-media-apps-like-xiaohongshu-pull-in-s-pore-users

Google. (n.d.). *Nudity & Sexual Content Policy.* YouTube Help. Retrieved September 20, 2024, from https://support.google.com/youtube/answer/2802002?hl=en&ref_topic=9282679

Gottfried, J., & Grieco, E. (2018, October 23). *Younger Americans are better than older Americans at telling factual news statements from opinions*. Pew Research Center. https://www.pewresearch.org/short-reads/2018/10/23/younger-americans-are-better-than-older-americans-at-telling-factual-news-statements-from-opinions/

Grant, J. I. (2021, June 07). *Australia's eSafety Commissioner targets abuse online as Covid-19 supercharges cyberbullying*. The Strategist. https://www.aspistrategist.org.au/australias-esafety-commissioner-targets-abuse-online-as-covid-19-supercharges-cyberbullying/

Hamm, J. A., Wylie, L. E., & Brank, E. M. (2016). Measuring Older Adult Confidence in the Courts and Law Enforcement. *Faculty Publications, Department of Psychology*, 682. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1682&context=psychfacpub

Hamzah, F. (2025, April 4). *Asian police forces nab more than 400 suspects in joint operation targeting online child abuse*. Channel NewsAsia. https://www.channelnewsasia.com/singapore/online-child-abuse-porn-arrests-cross-border-operation-5044101

Hern, A. (2020, March 13). *Covid-19 could cause permanent shift towards home working*. The Guardian. https://www.theguardian.com/technology/2020/mar/13/covid-19-could-cause-permanent-shift-towards-home-working

Howe, S. (2024, May 6). *Social Media Statistics in Singapore [Updated 2024]*. Meltwater. https://www.meltwater.com/en/blog/social-media-statistics-singapore

Infocomm Media Development Authority. (2025). *Online Safety Assessment Report 2024.* https://www.imda.gov.sg/-/media/imda/files/regulations-and-licensing/regulations/online-safety/online-safety-assessment-report-2024-designated-social-media-services.pdf

Infocomm Media Development Authority. (2023a). *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

Infocomm Media Development Authority. (2023b, July 17). *IMDA's Online Safety Code comes into effect* [Press release]. https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2023/imdas-online-safety-code-comes-into-effect

Infocomm Media Development Authority. (n.d.). *Code of Practice for Online Safety 2023.* https://www.imda.gov.sg/-/media/imda/files/regulations-and-licensing/regulations/codes-of-practice/codes-of-practice-media/guidelines-for-code-of-practice-for-online-safety.pdf

Intellectual Property Office of Singapore. (2025). *Copyright Resources - Factsheets on Copyright Act 2021*. https://www.ipos.gov.sg/about-ip/copyright/copyright-resources#3ec73b477055ea34a03d1e9b38f5ec32

Internet Matters. (n.d.). *Online safety leaflets and resources.* https://www.internetmatters.org/resources/esafety-leaflets-resources/

Lee, J-J. (2024, October 10). Cabinet approves bill revision to punish possessing, watching deepfake porn. *The Korea Herald*. https://www.koreaherald.com/article/3490966

LinkedIn. (n.d.). *Nudity and adult content*. LinkedIn Help. Retrieved September 18, 2024, from https://www.linkedin.com/help/linkedin/answer/a1338800

Middleton, K. (2024, June 1). *Jail time for those caught distributing deepfake porn under new Australian laws*. The Guardian. https://www.theguardian.com/australia-news/article/2024/jun/01/creating-or-sharing-deepfake-porn-without-consent-to-be-under-proposed-new-australian-laws

Minárik, T. (n.d.). *ASEAN to Focus on Cybersecurity Capacity- and Confidence-Building in 2017. The NATO Cooperative Cyber Defence Centre of Excellence*. https://ccdcoe.org/incyder-articles/asean-to-focus-on-cybersecurity-capacity-and-confidence-building-in-2017/

Ministry of Digital Development and Information. (2024, July 25). *MDDI Survey: Two Thirds of Respondents Encountered Harmful Online Content* [Press release]. https://www.mddi.gov.sg/media-centre/press-releases/mddi-survey-two-thirds-respondents-encountered-harmful-online-content/

Ministry of Digital Development and Information. (2023, October 17). *Survey by MCI finds that two thirds of Singapore users encountered harmful online content* [Press release]. https://www.mddi.gov.sg/media-centre/press-releases/survey-by-mci-on-harmful-online-content-encountered-by-sg-users/

Ministry of Digital Development and Information. (2022). *Sunlight Alliance for Action To Tackle Online Harms Detailed Sensing Poll Findings and Research Roadmap.* https://www.mddi.gov.sg/files/Press%20Releases%202022/sunlight%20afa%20sensing%20poll%20findings%20and%20research%20roadmap%20(2).pdf

Ministry of Law. (2024, November 22). *MinLaw and MDDI Seek Feedback on New Legislation and Measures to Enhance Online Safety and Better Protect Singaporeans from Harmful Online Content* [Press release]. https://www.mlaw.gov.sg/minlaw-and-mddi-seek-feedback-on-new-legislation-and-measures-to-enhance-online-safety/

Ministry of Justice. (2022). *How restorative justice works*. Retrieved May 27, 2025, from https://www.justice.govt.nz/courts/criminal/charged-with-a-crime/how-restorative-justice-works/

Online Criminal Harms Act 2023 (SG). https://sso.agc.gov.sg:5443/Act/OCHA2023?WholeDoc=1

Online Safety (Miscellaneous Amendments) Act 2022 (SG). https://sso.agc.gov.sg:5443/Acts-Supp/38-2022/Published/20221221?DocDate=20221221&WholeDoc=1

Online Safety and Media Regulation Act 2022 (Ireland). https://data.oireachtas.ie/ie/oireachtas/act/2022/41/eng/enacted/a4122.pdf

Organisation for Economic Co-operation and Development. (2021). *The role of online platforms in weathering the COVID-19 shock*. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/01/the-role-of-online-platforms-in-weathering-the-covid-19-shock_471ef11e/2a3b8434-en.pdf

Orme, B. (2005). *Accuracy of HB Estimation in MaxDiff Experiments*. Sawtooth Software, Inc. https://sawtoothsoftware.com/resources/technical-papers/accuracy-of-hb-estimation-in-maxdiff-experiments

Park, K., Ging, D., Murphy, S., & Mcgrath, C. (2023). *The impact of the use of social media on women and girls*. European Parliament. https://www.europarl.europa.eu/RegData/etudes/STUD/2023/743341/IPOL_STU(2023)743341_EN.pdf

Penal Code 1871 (SG). https://sso.agc.gov.sg/Act/PC1871?WholeDoc=1

Powell, A., Flynn, A., & Hindes, S. (2022). *Technology-facilitated abuse: National survey of Australian adults' experiences*. Australia's National Research Organisation for Women's Safety. https://anrows-2019.s3.ap-southeast-2.amazonaws.com/wp-content/uploads/2022/07/27172214/4AP.3-Flynn-TFa3-Survey-of-VS.pdf

Protection from Harassment Act 2014 (SG). https://sso.agc.gov.sg:5443/Act/PHA2014?WholeDoc=1

REACH. (2025, March 26). *Public Consultation on Enhancing Online Safety*. https://www.reach.gov.sg/latest-happenings/public-consultation-pages/2024/public-consultation-on-enhancing-online-safety#504caac6b77b97ee4c3cc8345083a68d

Restorative Justice Circle. (2016). *2016 Ipsos MORI poll summary.* https://restorativejustice.org.uk/sites/default/files/resources/files/2016%20Ipsos%20MORI%20poll%20summary.pdf

Revenge Porn Helpline. (n.d.). *About the Revenge Porn Helpline*. https://revengepornhelpline.org.uk/about-us/

Saucier, D. A., Miller, S. S., & O'Dea, C. J. (2025). Individual Differences in Masculine Honor Beliefs and Men's Perceptions of Insults Targeting Their Masculinity. *Journal of Interpersonal Violence*. https://doi.org/10.1177/08862605251329470

Sawtooth Software. (n.d.). *MaxDiff*. https://sawtoothsoftware.com/maxdiff

SG Her Empowerment. (2025). *404 Help Not Found: Lived Experiences of Online Harms Survivors.* https://api2.she.org.sg/uploads/404_Help_Not_Found_-_Lived_Experiences_of_Online_Harms_Survivors_-_SHE_Study_2025.pdf

SG Her Empowerment. (2023). *Study on online harms in Singapore 2023.* https://api2.she.org.sg/uploads/SHE_Report_on_Online_Harms_Study_Final.pdf

Soon, C., Goh, S. & Bala Krishnan, N. (2023, January). Study on Singaporeans and False Information Phase Two and Phase Three — Immunity and Intervention. *IPS Exchange Series No. 23*. Institute of Policy Studies. National University of Singapore. https://lkyspp.nus.edu.sg/docs/default-source/ips/ips-exchange-series-23.pdf

Stevens, F., Enock, F. E., Sippy, T., Bright, J., Cross, M., Johansson, P., Wajcman, J., & Margetts, H. Z. (2024). Understanding gender differences in experiences and concerns surrounding online harms: A nationally representative survey of UK adults. Alan Turing Institute. https://www.turing.ac.uk/sites/default/files/2024-03/understanding_gender_differences_in_experiences_and_concerns_surrounding_online_harms_-_a_nationally_representative_survey_of_uk_adults.pdf

Szeto, S., Au, A. K. Y., & Cheng, S. K. L. (2024). Support from Social Media during the COVID-19 Pandemic: A Systematic Review. *Behavioral Sciences, 14*(9), 759. https://doi.org/10.3390/bs14090759

Tan, S. (2024, February 7). *Match, Chat, Love: Examining the popularity and usage of dating apps in Singapore*. YouGov. https://sg.yougov.com/consumer/articles/48571-match-chat-love-examining-the-popularity-and-usage-of-dating-apps-in-singapore

The Asia Foundation. (2023). *Study on public attitude toward restorative justice implementation in Indonesia*. https://asiafoundation.org/wp-content/uploads/2023/08/EN_Executive-Summary-Study-on-Public-Attitude-Toward-Restorative-Justice-Implementation-in-Indonesia.pdf

Venkataramakrishnan, S., & Squirrell, T. (2024). *The 'Manosphere'.* Institute for Strategic Dialogue. https://www.isdglobal.org/explainers/the-manosphere-explainer/

Wang, M-J., Yogeeswaran, K., Andrews, N. P., Hawi, D. R., & Sibley, C. G. (2019). How Common Is Cyberbullying Among Adults? Exploring Gender, Ethnic, and Age Differences in the Prevalence of Cyberbullying. *Cyberpsychology, Behavior, and Social Networking, 22*(11), 736-741. https://doi.org/10.1089/cyber.2019.0146

Wenzel, M., Okimoto, T. G., Feather, N. T., & Platow, M. J. (2008). Retributive and restorative justice. *Law and Human Behavior, 32*(5), 375–389. https://doi.org/10.1007/s10979-007-9116-6

We Are Social., & Meltwater. (2024). Digital 2024: 5 billion social media users. https://wearesocial.com/sg/blog/2024/01/digital-2024-5-billion-social-media-users/

World Economic Forum. (2023). *Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms*. https://www3.weforum.org/docs/WEF_Typology_of_Online_Harms_2023.pdf

Zizzola, A. (n.d.). *Restorative Justice Responses to Cyber Harm.* European Forum for Restorative Justice. https://www.euforumrj.org/restorative-justice-responses-cyber-harm

APPENDICES

Appendix A: Phase One methodology — Landscape review

To ensure a comprehensive account of online harms impacting communities and individuals, we conducted a landscape scan that examined a broad range of online safety regulations from multiple jurisdictions between June 2024 to October 2024. We shortlisted 21 key online safety regulations, including laws and codes of practice, from nine jurisdictions that specifically address online harms or include provisions targeting online harms activities. We reviewed regulations including recent online safety regulations from Singapore, UK and Ireland, legislations from Asia that cover the usage of computers and transnational regulations that govern data privacy and digital services. The regulations cover online harms across all ages, instead of only those applicable to children and minors.

We also examined the publicly accessible community guidelines and standards of 21 social media platforms that are commonly used by Singaporeans. These platforms were selected based on two key criteria: the volume of Singaporean users and the average time spent by Singaporean users on each platform. Our review also included the top three messaging apps by number of users in Singapore, i.e., WhatsApp, Telegram and WeChat.[13]

[13] Howe, S. (2024, May 6). *Social Media Statistics in Singapore [Updated 2024]*. Meltwater. https://www.meltwater.com/en/blog/social-media-statistics-singapore

In addition to the most popular platforms used by Singaporeans, we also included a selection of less mainstream platforms that emphasised community-building and user-led moderation such as Reddit, Discord, Twitch, Lemon8 and Mastodon. These platforms were included due to their unique community dynamics and moderation structures, which may influence safety and content governance in ways distinct from the more conventional platforms.

We also reviewed Rednote (also known as "Xiaohongshu" in Chinese) due to its growing user base among Singaporeans. The platform has an estimated 600,000 users in Singapore, making it a relevant inclusion outside the top social media platforms locally.[14] Additionally, we included the top four dating platforms used by Singaporeans given their growing popularity and the unique safety considerations they present.[15] Tables A1 and A2 present the list of regulations and platforms reviewed for our paper.

[14] Goh, Y. H. (2024, April 24). "Black hole" of alluring content: Chinese social media apps like Xiaohongshu pull in S'pore users. *The Straits Times*. https://www.straitstimes.com/asia/east-asia/black-hole-of-alluring-content-chinese-social-media-apps-like-xiaohongshu-pull-in-s-pore-users; Howe, S. (2024, May 6). *Social Media Statistics in Singapore [Updated 2024]*. Meltwater. https://www.meltwater.com/en/blog/social-media-statistics-singapore

[15] Tan, S. (2024, February 7). *Match, Chat, Love: Examining the popularity and usage of dating apps in Singapore*. YouGov. https://sg.yougov.com/consumer/articles/48571-match-chat-love-examining-the-popularity-and-usage-of-dating-apps-in-singapore

Table A1: Regulations reviewed

Regulation	Origin
Singapore Computer Misuse Act 1993	Singapore
Singapore Protection from Harassment Act 2014	
Singapore Protection from Online Falsehoods and Manipulation Act 2019	
Singapore Foreign Interference Countermeasures Act 2021	
Singapore Online Safety (Miscellaneous Amendments) Act 2022	
Singapore Online Criminal Harms Act 2023	
Singapore Code of Practice for Online Safety 2023	
EU General Data Protection Regulation 2016	European Union
EU Digital Services Act 2022	
UK Online Safety Act 2023	United Kingdom
Ireland Online Safety and Media Regulation Act 2022	Ireland
Germany Network Enforcement Act 2017	Germany
Australia Online Safety Act 2021	Australia
Australia eSafety Commissioner Basic Online Safety Expectations Regulatory Guidance 2024	
Australian Code of Practice on Disinformation and Misinformation 2022	
China Cybersecurity Law 2017	China
China Guiding Opinions on Punishing Cyberviolence Violations and Crimes in Accordance with Law 2023	
China Regulations on the Protection of Minors Online 2023	
India Information Technology Act 2000	India
Malaysia Computer Crimes Act 1997	Malaysia
Malaysia Communications and Multimedia Act 1998	

Table A2: Platforms reviewed

Type of platform	Platform
More mainstream social media platforms in Singapore	TikTok (Bytedance)
	Facebook (Meta)
	Instagram (Meta)
	YouTube (Google)
	X
	LinkedIn
Most popular instant messaging platforms in Singapore	Telegram
	WhatsApp (Meta)
	WeChat
Less mainstream social media platforms in Singapore	Snapchat
	Reddit
	Discord
	Twitch
	Omegle
	Mastodon
	Lemon8 SG (Bytedance)
	Rednote (“Xiaohongshu”)
Dating platforms	Tinder (Match Group)
	OKC (Match Group)
	CMB
	Bumble

Beside regulations and community guidelines, we also conducted a scan of academic literature on online harms to identify online harms that are not covered by regulations and platforms. In all, we reviewed 47 studies conducted by academics, think tanks and non-government organisations, with a focus on Singapore and Southeast Asia. We only reviewed studies that examined an array of online harms and did not include studies that focused specifically on one type of online harm.

Through our review, we found 12 harms that were not included in the World Economic Forum (WEF) typology and 10 harms not included in IMDA's Code of Practice for Online Safety (CoP). See Table A3 for the list of the harms covered.

Table A3: List of online harms covered by the WEF Typology and IMDA CoP

Harms	WEF Typology	IMDA CoP
Catfishing*	✓	✓
Child sexual abuse material*	✓	✓
Child sexual exploitation material*	✓	✓
Content that incites, promotes or facilitates violence*	✓	✓
Content that incites, promotes or instructs dangerous physical behaviour*	✓	✓
Content that praises, promotes, glorifies or supports extremist organisations or individuals*	✓	✓
Developmentally inappropriate content*	✓	✓
Disinformation and misinformation*	✓	✓
Doxxing*	✓	✓
Grooming for sexual abuse*	✓	✓
Hate speech*	✓	✓
Image-based abuse*	✓	✓
Impersonation*	✓	✓
Material that promotes suicide, self-harm and disordered eating*	✓	✓
Online bullying and harassment*	✓	✓
Phishing*	✓	✓

Pro-terror material*	✓	✓
Recruitment and radicalisation*	✓	✓
Scams*	✓	✓
Sexual extortion*	✓	✓
Technology-facilitated abuse*	✓	✓
Technology-facilitated gender-based violence*	✓	✓
Violent graphic content*	✓	✓
Abuse of reporting tools		
Adult nudity and sexual content	✓ (under Developmentally Inappropriate content)	✓
Content that impacts legal process or ongoing legal proceedings		
Sale or supply of regulated, restricted or illegal goods and services		
Election interference	✓ (under Disinformation and Misinformation)	
Evasion of enforcement		
Forgery		
Foreign interference		
Incitement of criminal activity		✓
Promotion of illegal activities or behaviour		✓
Sexual solicitation, exploitation or prostitution		
Spam		
Threats to public safety		✓ (under Content Endangering Public Health)
Unauthorised access or use		

* From the WEF Typology of Online Harms

Appendix B: Online harms coverage

Based on our review, we identified 20 online harms covered by both regulations and the community guidelines of social media platforms. Table B1 lists the harms in order of coverage by the 42 regulations and guidelines.

Table B1: Harms covered by both regulations and community guidelines

	Online harms covered by both regulations and community guidelines	Coverage (42)
1.	Graphic violence and threats of violence	30
2.	Hate speech and hateful behaviour	30
3.	Harassment and cyberbullying	30
4.	Adult nudity and sexual content	29
5.	Doxxing	27
6.	Child sexual exploitation and/or abuse	27
7.	Misinformation and disinformation	27
8.	Promotion or glorification of self-harm	25
9.	Promotion or support of extremist of terrorist organisations and individuals	26
10	Sale or supply of regulated, restricted or illegal goods and services	22
11	Cheating, fraud and scams	20
12	Sextortion	20
13	Impersonation and catfishing	21
14	Unauthorised access or use of computers or platform	16
15	Sexual solicitation, exploitation or prostitution of adults	16
16	Intellectual property or copyright infringement	16
17	Election interference	15
18	Human exploitation and trafficking	15
19	Fake engagement	14
20	Promotion of dangerous physical activities or behaviour	12

Our review also revealed four harms that are only covered by social media platforms (see Table B2) and another four harms that are only covered by regulations (see Table B3).

Table B2: Harms covered by platforms only

	Online harms covered only by social media platforms	Coverage (21)
1.	Spam	16
2.	Promotion of illegal activities or behaviours	12
3.	Abuse of reporting tools	9
4.	Evasion of enforcement	9

Table B3: Harms covered by regulations only

	Online harms covered only by regulations	Coverage (21)
1.	Incitement of criminal activity	6
2.	Threats to public safety	6
3.	Forgery	4
4.	Content that impacts ongoing legal proceedings or processes	3

Appendix C: Phase Two methodology — Focus group discussions

Participants were recruited from a database maintained by the Institute of Policy Studies Social Lab (IPSSL), consisting of individuals who had previously agreed to be contacted for research studies. Additional participants were recruited through researchers' professional and personal networks. To capture diverse perspectives, focus group discussions (FGDs) were structured based on age groups, gender, and familial and occupational roles. Table C1 presents a summary of the groups formed and their rationale.

Table C1: Summary and rationale for focus groups

Groups	Target Segment	Rationale
1 & 2	Females (17–24 years old) and Females (25–34 years old)	Existing research indicates that youths are the most affected groups. These two age groups are in different life stages, have different media and social routines, and would likely have different perspectives of online harms.
3 & 4	Males (17–24 years old) and Males (25–34 years old)	Recent studies suggest that males encounter different online harms from females. Groups 1 through 4 were designed to explore gendered perspectives of online harms and their severity.
5 & 6	Females (35 years old & above), and Males (35 years old and above)	Likewise, these groups have been selected to examine age-related differences for older males and females.
7 & 8	Parents, Educators and Social Service Professionals	These groups play key roles in safeguarding young people from online harms. Social service professionals such as counsellors and youth workers triage online harms and offer important perspectives on the impact of online harms and the effectiveness of interventions.

9 & 10	Industry Partners and Safety Tech Providers	Industry partners include individuals such as academics or professionals who assess how public policies are implemented and whether they are functioning as intended. Safety tech providers refer to organisations that focus on developing technologies or solutions to create a safer digital space and online experiences. These groups offer unique perspectives from the supply side of online platforms and safety tech, providing insights into technical and industry definitions of online harms.

In total, 10 FGDs were conducted in-person between September and December 2024, involving 79 participants. See Table C2 for the demographic breakdown of participants.

Table C2: Demographics of focus group participants

Category	**Number of individuals**
Group 1 – Females (17–24 years old)	8
Group 2 – Females (25–34 years old)	8
Group 3 – Males (1 –24 years old)	10
Group 4 – Males (25–34 years old)	10
Group 5 – Females (35 years old & above)	8
Group 6 – Males (35 years old & above)	8
Group 7 – Parents	6
Group 8 – Educators & Social Service Professionals	8
Group 9 – Industry Partners	7
Group 10 – Safety Tech Providers	6

Appendix D: Ranking activity (FGD)

To understand how different groups define and prioritise what they consider to be harmful, participants engaged in a ranking exercise using printed cards that listed 12 predefined types of online harms (see Table D1). Each participant first ranked these harms individually before engaging in a group discussion to develop a collaborative ranking. This process encouraged deliberation, allowing participants to explain their own rankings and reconcile differing perspectives. Some groups identified additional harms, expanding their rankings beyond the original 12 categories.

Table D1: List of online harms for ranking activity

Categories	Examples
Child Sexual Exploitation and Abuse	Child pornography, grooming
Cyberbullying/Harassment	Sending abusive texts and emails, threats or derogatory or dehumanizing statements or references related to the target person's characteristics, posting or threatening to post personal information of the target person online (doxxing), online stalking
Hateful online material	Slurs, hate speech against a target group because of certain characteristics, e.g., religion, ethnicity, sex, sexual orientation
Violent/terrorism-related content	Images of gore, sadistically celebrating violent images, abuse or torture on persons or animals, content providing instruction on terrorist acts, threats of violence
Promotion of dangerous behaviours	Content promoting suicide / self-harm, images or posts of pro-eating disorders content
False or misleading online material	Misinformation, disinformation, making false statements of fact about an individual that cause harm, content that misleads people about when, where, or how to participate in an election or other civic processes
Misuse of inauthentic material	Malicious synthetic / manipulated content**,** AI-generated deepfakes to depict a person without the person's consent
Sexual content	Depiction of explicit sexual activities, deviant sexual behaviour (e.g., paedophilia, incest), voyeuristic or intimate image or recording distributed without consent, sextortion
Content that could endanger public health	Discouraging vaccination or following health safety measures
Content supporting vice and organised crime	Scams, sale of illegal goods such as vapes and drugs, encouraging crime, sexual solicitation, exploitation or prostitution of adults, human trafficking
Impersonation	Identity theft (e.g., posing as someone else online, by creating an online account using someone else's name or photos)
Online statements instigating disproportionate harm	Cancel campaign to punish an individual for what he/she may have said or done, making a statement that would lower the reputation of an individual (e.g., cause loss of employment

Appendix E: Phase Two methodology — In-depth interviews

The in-depth interviews (IDIs) elicited a more nuanced picture of victim's journey and processes that they have taken when they are seeking recourse for the online harms they encountered. Our collaborator, SG Her Empowerment, conducted IDIs with 20 participants — 12 victims of online harms and eight supporters between October 2024 and January 2025.

Victims were recruited from SHECARES@SCWO, Singapore's first dedicated support centre focused on assisting victims of online harms. The centre provides counselling, legal assistance, and support in filing reports to police or applications to the Protection from Harassment Courts (PHC). Additionally, it collaborates with internet platforms to facilitate the removal of harmful online content. Only closed cases at the centre were recruited for the study to prevent interruption in therapeutic interventions individuals were receiving at the centre. See Table E1 for victims' demographic breakdown and case IDs.

Table E1: Victims' demographic data and case IDs

Type of Harm	**Count**	**Cases**
Cyberbullying/Harassment	10	QV01, QV02, QV03, QV04, QV05, QV07 QV08, QV09, QV10, QV12
IBSA	4	QV03, QV04, QV06, QV11
Sexual Harassment	3	QV05, QV09, QV08
Doxxing	5	QV01, QV07, QV08, QV09, QV12
Secondary Harm[16]		
Impersonation	3	QV05, QV07, QV08

[16] The two typologies of harm were considered secondary harms, as they extend and amplify the effects of the primary harm instead of occurring in isolation.

Cyberstalking	8	QV02, QV03, QV04, QV05, QV07, QV08, QV09, QV10
Age	**Count**	**Cases**
20–30 years old	6	QV01, QV03, QV04, QV06, QV08, QV10, QV11
31–40 years old	3	QV02, QV05, QV09, QV12
> 40 years old	1	QV07
Gender	**Count**	**Cases**
Male	2	QV04, QV12
Female	10	QV01, QV02, QV03, QV05, QV06, QV07, QV08, QV09, QV10, QV11

Supporters were recruited from a range of stakeholders in the support ecosystem — lawyers, social service professionals (including caseworkers and counsellors) and teachers. They shared approaches to and their experience in triaging online harms. See Table E2 for victims' demographic breakdown and case IDs.

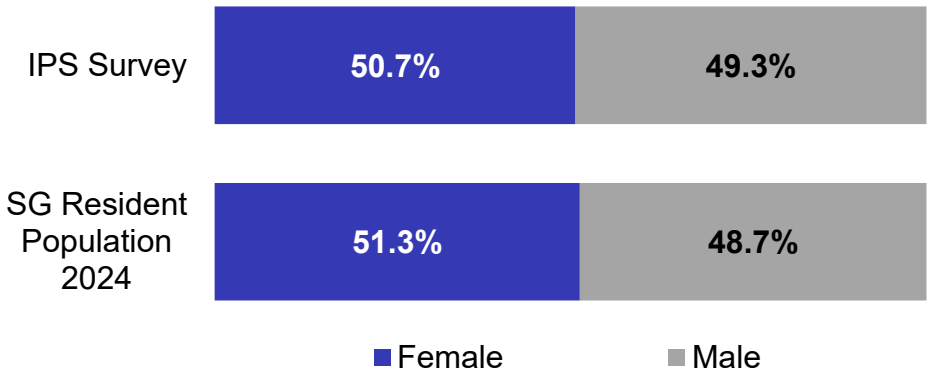
Table E2: Supporters' demographic data and case IDs

Occupation	**Count**	**Case ID**
Lawyer	2	QS01, QS06
Social service professional	3	QS02, QS03, QS04
Online harm centre counsellor	1	QS05
Teacher	2	QS07, QS08

Appendix F: Phase Three methodology — Online validation survey

The survey data was collected by IPS Social Lab using Toluna's online consumer panel. Only Singapore Citizens and Permanent Residents aged 17 years old and above were eligible to participate in the online survey. The self-administered survey was available in English only and the average duration was 20 minutes. Prior to the full fieldwork, we conducted a pilot survey with around 60 respondents. No issues surfaced during the pilot survey and the fieldwork took place in March 2025. The final sample comprised 600 respondents after quality checks.[17] Overall, our final sample closely matched the gender distribution of Singapore's resident population (see Figure F1).

Figure F1: Representation by gender

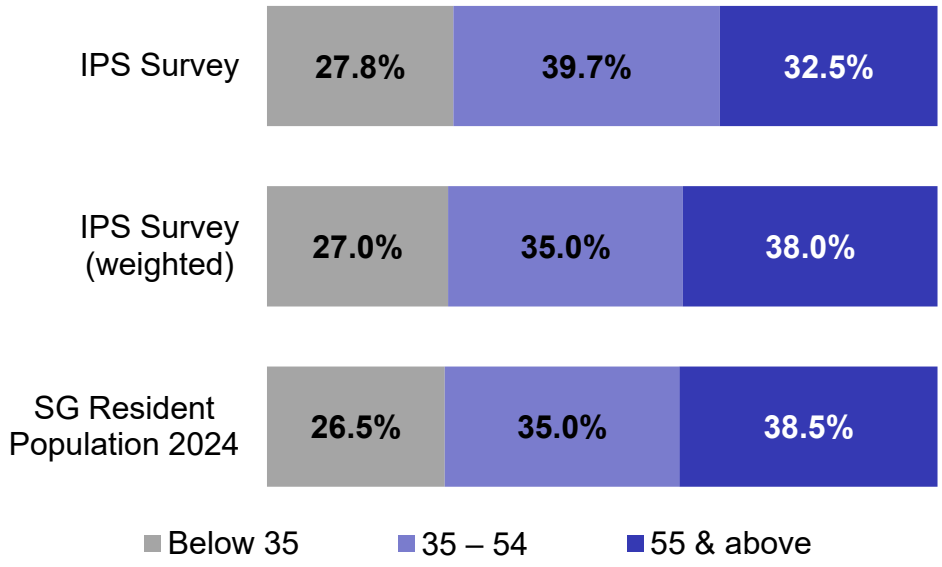


While the study aimed to obtain a sample that approximates the demographic traits of Singapore's resident population, the oldest age group was slightly under-represented given the online nature of the data collection. For our

[17] Data from respondents who failed the attention check question or had straight-lined responses were removed for analyses purposes.

analysis, age was grouped into three segments: youth (below 35 years old), middle-aged (35 to 54 years old), and older participants (55 years old and above). Statistical weights were applied to ensure that the survey sample matched more closely to the resident population's age distribution in these three categories for 2024 (See Figure F2).

Figure F2: Representation by age group



Appendix G: Maximum differences scaling (MaxDiff)

The MaxDiff (Maximum Difference Scaling), also known as best-worst scaling, is a quantitative method used to determine preferences or priorities among a set of items. It involves displaying sets of 3-6 items to respondents and asking them to evaluate the best and worst (or most preferred vs. least preferred) choices in each set.[18] By analysing responses across multiple sets, MaxDiff produces a ranking of all items based on their relative importance.

In this study, MaxDiff was used to assess the relative severity of various online harms. This method was chosen over more traditional approaches such as Likert scales or ranking questions as it allowed us to obtain greater distinction between the different harms we were testing. By forcing respondents to make trade-offs between items, MaxDiff produces more discriminating and meaningful results.

A total of 16 online harms were evaluated (see Appendix H). Each respondent was presented with 12 questions, and was shown a randomised set of four online harms for each question. For each set, respondents were instructed to indicate the harm they perceived as most severe and the one they perceived as least severe. Figure G1 shows a screenshot of the survey user interface.

[18] Sawtooth Software. (n.d.). *MaxDiff*. https://sawtoothsoftware.com/maxdiff

Figure G1: Screenshot of survey user interface for MaxDiff questions

Most severe online harm		Least severe online harm
○	Statements **affecting reputation** i.e., making a statement, whether fact or opinion, that would lower the reputation of a person	○
○	**Misinformation about health risks** such as unproven treatments	○
○	**Promotion of dangerous behaviours** such as self-harm, suicide or eating disorders	○
○	**Sexual content** depicting **sexual activities or sexual behaviours**, which were **recorded consensually**.	○

Each harm appeared exactly three times across the 12 questions, ensuring a balanced design where all items were evaluated equally. In total, each harm appeared 1,800 times for 600 respondents. This approach allowed for a robust estimation of the relative severity attributed to each harm by the respondents.

The MaxDiff analysis generated three different outputs. The first was a relative severity score calculated for each online harm. This score was derived by subtracting the total number of times the harm was selected as the most severe from the total number of times it was selected as the least severe.

The second output was preference shares. Preference shares go a step further from the best-worst analysis to account for the total preference that each harm received across all respondents, taking into account the context (which items were shown together) and the trade-offs respondents made across sets. These shares were estimated using Hierarchical Bayes (HB) modelling, which leverages a multinomial logit framework to derive individual-level utility scores. HB is widely regarded as the gold standard in MaxDiff analysis, as it captures heterogeneity across respondents and produces stable, high-quality estimates

that can be meaningfully aggregated.[19] We generated 100-sum scores based on preference shares to provide a model based on the proportion of relative importance that adds up to 100 per cent.

Finally, a comparison table for severity rankings of online harms was generated using preference shares to calculate head-to-head comparisons between harms. This table compares each harm directly against every other harm, providing a detailed view of how often one harm was rated as more severe than another across all possible pairings.

Together, these outputs offer a comprehensive understanding of how respondents perceived the various online harms included in the study.

[19] Orme, B. (2005). *Accuracy of HB Estimation in MaxDiff Experiments*. Sawtooth Software, Inc. https://sawtoothsoftware.com/resources/technical-papers/accuracy-of-hb-estimation-in-maxdiff-experiments

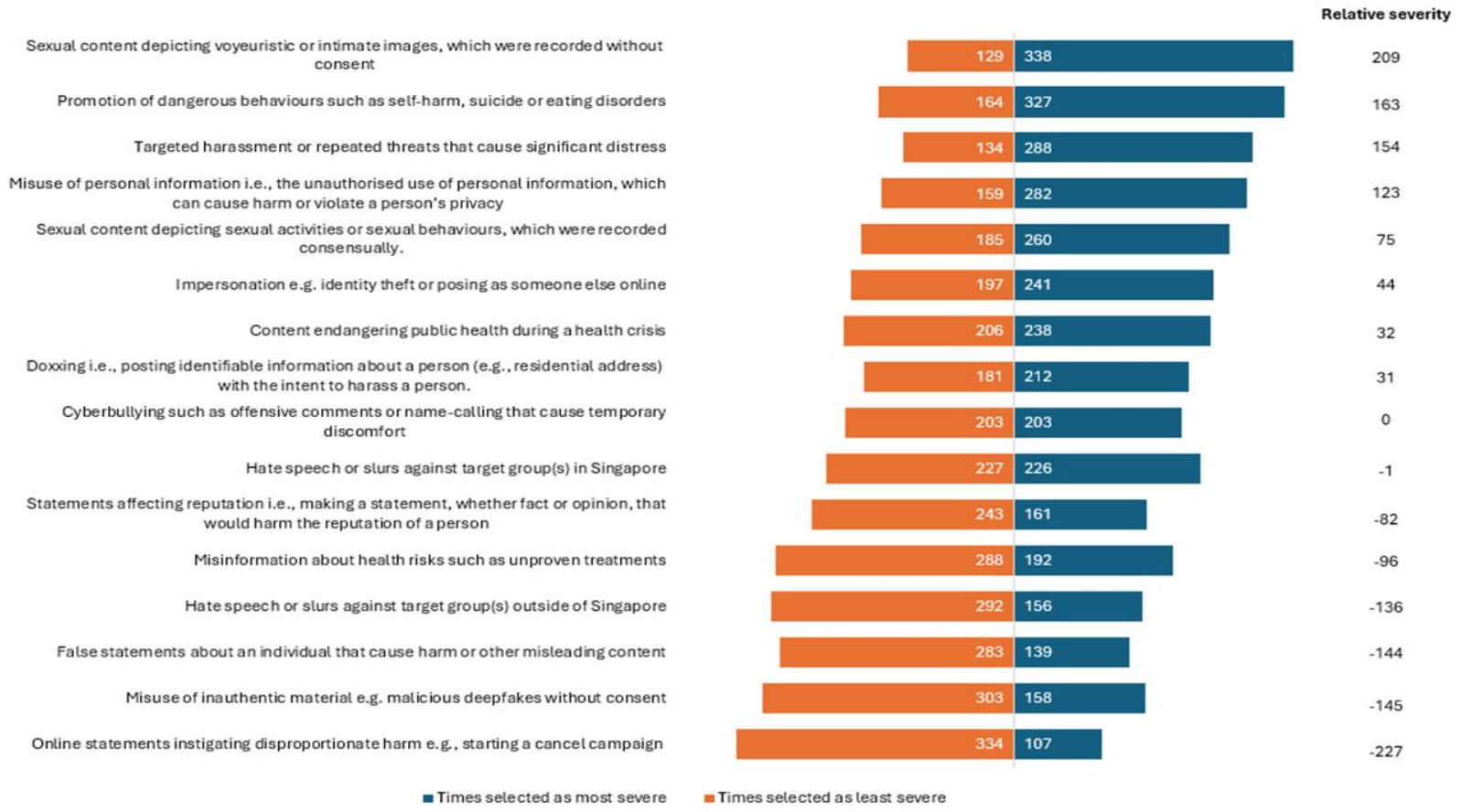
Appendix H: List of online harms for ranking (Survey)

Types of online harms
Sexual content depicting voyeuristic or intimate images, which were **recorded without consent**
Promotion of dangerous behaviours such as self-harm, suicide or eating disorders
Sexual content depicting sexual activities or sexual behaviours, which were **recorded consensually**
Misuse of personal information, i.e., the unauthorised use of personal information, which can cause harm or violate a person's privacy
Targeted harassment or repeated threats that cause **significant distress**
Doxxing, i.e., posting identifiable information about a person (e.g., residential address) with the intent to harass a person
Impersonation, e.g., identity theft or posing as someone else online
Content endangering public health **during a health crisis**
Cyberbullying such as offensive comments or name-calling that cause **temporary discomfort**
Hate speech or slurs against target group(s) **in Singapore**
Statements affecting reputation, i.e., making a statement, whether fact or opinion, that would harm the reputation of a person
Misinformation about health risks such as unproven treatments
Misuse of inauthentic material, e.g., malicious deepfakes without consent
Hate speech or slurs against target group(s) **outside of Singapore**
False statements about an individual that cause harm or other misleading content
Online statements instigating disproportionate harm, e.g., starting a cancel campaign

Note: Harms that are colour-coded was compared against one another for hypotheses testing.

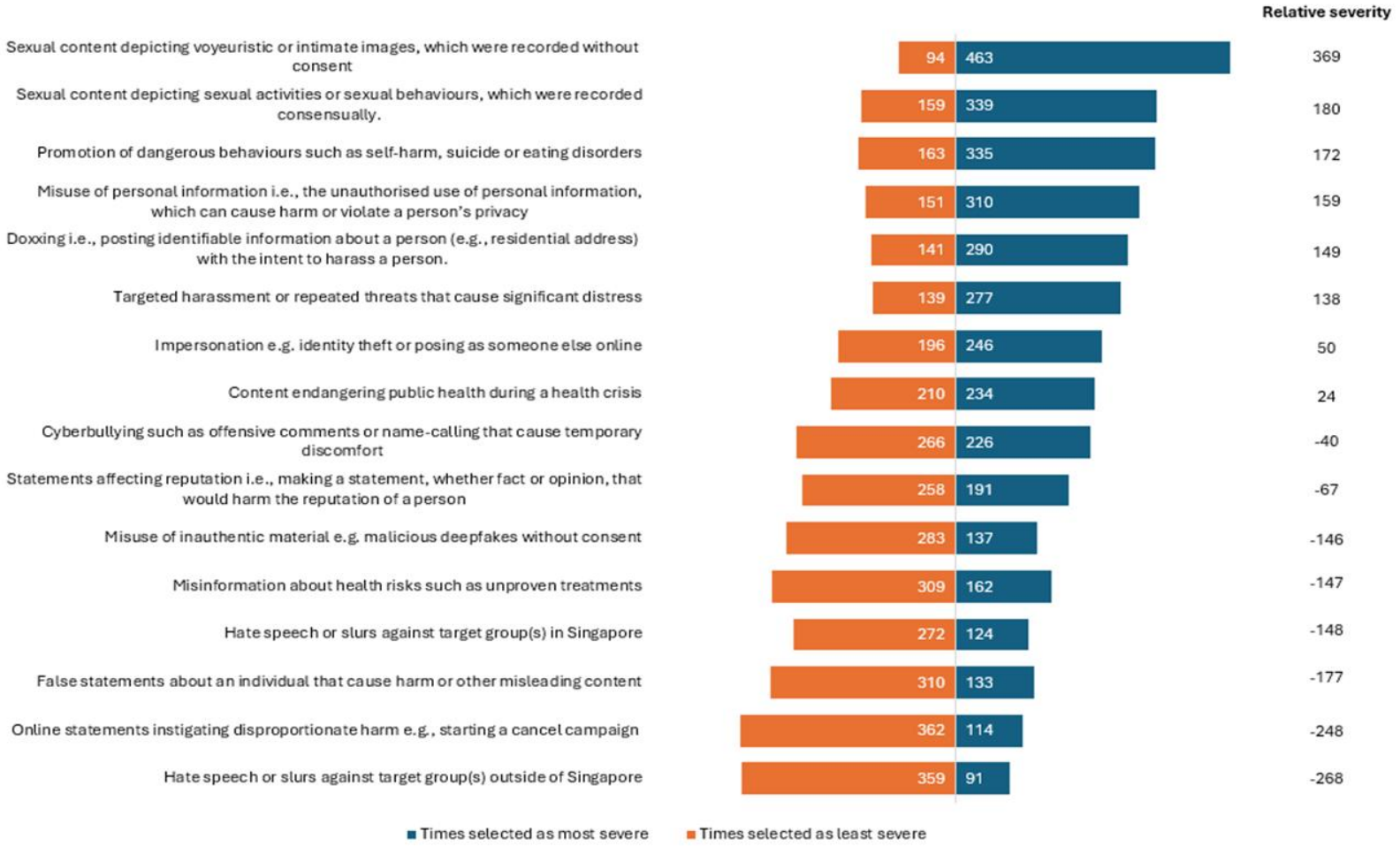
Appendix I: Times selected as most severe and least severe

Figure I1: Online harms severity rankings for males



Note: Analyses were performed on unweighted survey data.

Figure I2: Online harms severity rankings for females

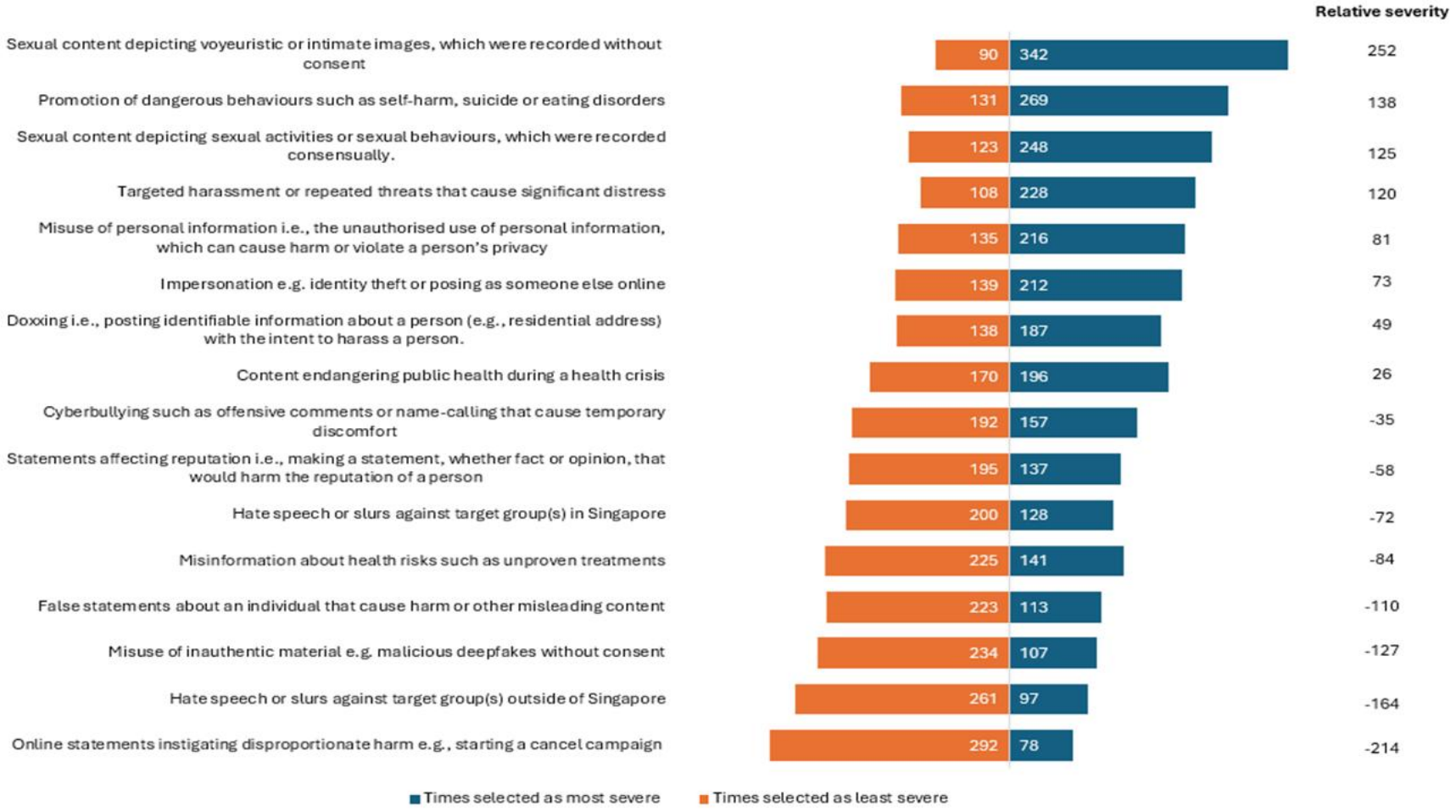


Note: Analyses were performed on unweighted survey data.

Figure I3: Online harms severity rankings for respondents below 35 years old

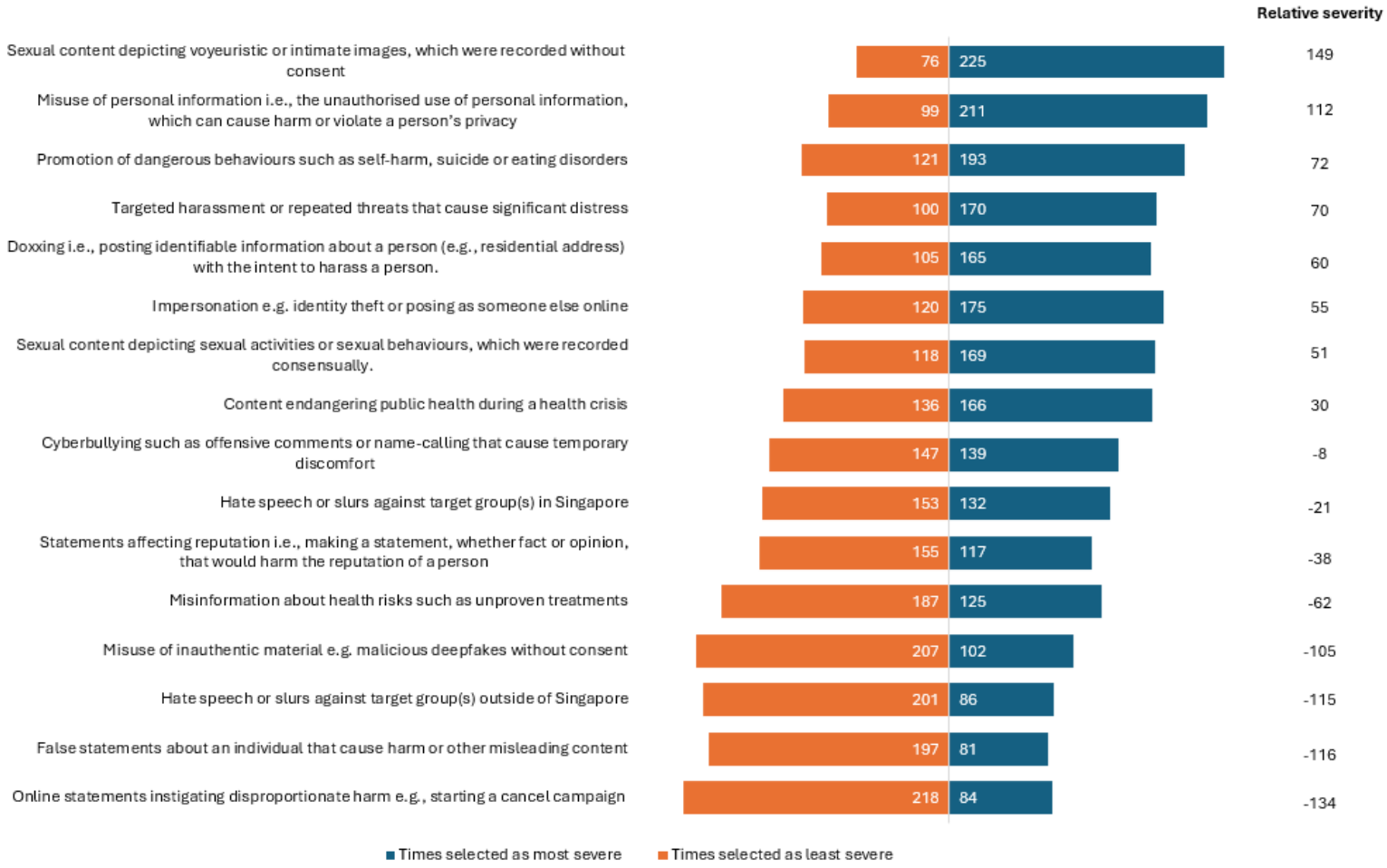
Note: Analyses were performed on unweighted survey data.

Figure I4: Online harms severity rankings for respondents aged 35–54 years old



Note: Analyses were performed on unweighted survey data.

Figure I5: Online harms severity rankings for respondents aged 55 years old and above



Note: Analyses were performed on unweighted survey data.

Appendix J: Factors influencing severity perceptions, by gender

	Males			Females				
Stakeholders	**M**	**SD**	**Top 2**	**M**	**SD**	**Top 2**	**Grand M**	**p**
How severe the harm is to individuals	4.04	0.90	75.1%	3.98	0.93	71.8%	4.01	n.s.
How vulnerable or susceptible the potential victims are	4.05	0.85	74.9%	3.96	0.97	73.8%	4.00	n.s.
Whether there are existing measures to address the harm (e.g., laws and safety features on social media)	4.02	0.84	74.0%	3.91	1.00	68.5%	3.97	n.s.
How many people are likely to be affected by the harm	4.06	0.87	74.0%	3.86	0.98	67.6%	3.96	.01*
How severe the harm is to the society	4.00	0.92	71.3%	3.88	0.95	68.4%	3.94	n.s.
How likely the harm will happen to people who are like me	3.83	0.95	68.4%	3.81	0.95	64.0%	3.82	n.s.

*p<.05

n.s. indicate non-significant findings

Appendix K: Stakeholders' responsibility, by gender

	Males			Females				
Stakeholders	**M**	**SD**	**Top 2**	**M**	**SD**	**Top 2**	**Grand M**	**p**
Singapore government	4.12	0.90	76.6%	4.16	0.95	77.5%	4.14	n.s.
Tech companies	4.14	0.92	76.5%	4.14	0.85	77.5%	4.14	n.s.
Users ourselves	4.13	0.93	75.5%	4.14	0.88	75.7%	4.14	n.s.
Parents	4.05	0.91	73.1%	4.17	0.84	79.9%	4.11	n.s.
Schools	4.07	0.90	75.5%	4.13	0.90	75.9%	4.10	n.s.
Social service agencies	3.79	0.95	62.4%	3.98	0.87	72.8%	3.89	.01*

*p<.05

n.s. indicate non-significant findings

Appendix L: Post hoc analyses for stakeholders' responsibility, by age

	<35 years old (A)			35-54 years old (B)			55 years old & above (C)						
Stakeholders	**M**	**SD**	**Top 2**	**M**	**SD**	**Top 2**	**M**	**SD**	**Top 2**	**Grand M**	**p (A vs B)**	**p (B vs C)**	**p (A vs C)**
Singapore government	3.95	0.99	69.4%	4.08	0.94	75.2%	4.32	0.82	84.1%	4.14	n.s.	.02*	<.001***
Tech companies	3.94	0.92	73.0%	4.11	0.92	74.4%	4.30	0.79	82.6%	4.14	n.s.	n.s.	<.001***
Schools	4.01	0.89	73.0%	4.02	0.96	73.1%	4.24	0.82	80.0%	4.10	n.s.	.02*	.03*
Social service agencies	3.77	0.96	65.3%	3.79	0.94	62.2%	4.06	0.84	74.3%	3.89	n.s.	.006**	.008**

* p<.05, ** p<.01, *** p<.001

n.s. indicate non-significant findings.

Appendix M: Post hoc analyses for remedies by social media services

	<35 years old (A)			35-54 years old (B)			55 years old & above (C)						
Remedies	**M**	**SD**	**Top 2**	**M**	**SD**	**Top 2**	**M**	**SD**	**Top 2**	**Grand M**	**p (A vs B)**	**p (B vs C)**	**p (A vs C)**
When social media services remove harmful content and accounts more quickly	4.04	0.97	71.8%	4.07	0.98	74.4%	4.29	0.82	83.1%	4.14	n.s.	.03*	.02*
When social media services implement safety by design (i.e., minimisation of risks and harms to users when developing products and services)	3.76	0.90	61.7%	3.91	0.91	69.3%	4.13	0.83	79.5%	3.96	n.s.	.02*	<.001***
When social media services put in place additional safety measures to protect children (e.g., age verification measures)	3.82	0.99	64.0%	3.88	0.96	68.1%	4.19	0.86	82.1%	3.98	n.s.	.001**	<.001***
When social media services improve user reporting tools and processes	3.90	0.95	66.4%	3.80	0.99	65.1%	4.06	0.80	78.4%	3.93	n.s.	.01*	n.s.
When social media services publish transparency accountability reports	3.77	0.98	62.9%	3.73	1.04	60.1%	4.02	0.90	72.8%	3.85	n.s.	.007**	.03*

*p<.05, ** p<.01, *** p<.001

n.s. indicate non-significant findings.

Appendix N: Post hoc analyses for public education remedies

	<35 years old (A)			35-54 years old (B)			55 years old & above (C)						
Remedies	**M**	**SD**	**Top 2**	**M**	**SD**	**Top 2**	**M**	**SD**	**Top 2**	**Grand M**	**p (A vs B)**	**p (B vs C)**	**p (A vs C)**
When public education efforts and campaigns tailor content to my age group and life stage	3.93	0.95	70.0%	3.74	0.98	61.3%	4.02	0.90	71.8%	3.90	n.s.	.008**	n.s.
When public education efforts and campaigns tell me what I should do by providing step-by-step processes on actions to take when I encounter online harms, or if I think I have encountered online harms	3.74	0.93	62.3%	3.81	1.00	63.9%	4.10	0.82	78.5%	3.90	n.s.	.003**	<.001***

*p<.05, ** p<.01, *** p<.001

n.s. indicate non-significant findings

Appendix O: Post hoc analyses for legislative remedies

	<35 years old (A)			35-54 years old (B)			55 years old & above (C)						
Remedies	**M**	**SD**	**Top 2**	**M**	**SD**	**Top 2**	**M**	**SD**	**Top 2**	**Grand M**	**p (A vs B)**	**p (B vs C)**	**p (A vs C)**
When legislation/law holds perpetrators accountable for their actions	4.12	0.98	76.0%	4.14	0.93	76.1%	4.36	0.76	84.6%	4.12	n.s.	.02*	.02*
When legislation/law can allow for take down of harmful online content	4.05	0.95	76.6%	4.14	0.93	76.9%	4.19	0.90	78.4%	4.05	n.s.	n.s.	n.s.
When legislation/law empowers victims to seek relief and financial damages for the harm suffered	3.97	0.98	73.0%	3.97	0.92	69.3%	4.22	0.84	81.5%	3.97	n.s.	.01*	.02*
When legislation/law is accompanied by public education on the protections conferred by existing laws and measures	3.88	0.87	70.1%	3.88	0.94	68.5%	4.11	0.80	79.0%	3.88	n.s.	.02*	.03*

*p<.05, ** p<.01, *** p<.001

n.s. indicate non-significant findings.

Appendix P: Independent t-test for online harms experience, by gender

Males			Females				
M	**SD**	**Top 2**	**M**	**SD**	**Top 2**	**Grand mean**	**p**
2.54	1.70	34.6%	2.21	1.50	38.1%	2.37	.01*

*p<.05

Appendix Q: Online harms experience, by age

<35 years old (A)			35-54 years old (B)			55 years old & above (C)		
M	**SD**	**Top 2**	**M**	**SD**	**Top 2**	**M**	**SD**	**Top 2**
2.63	1.61	41.9%	2.59	1.83	33.7%	1.98	1.29	34.9%

Grand M	p (A vs B)	p (B vs C)	p (A vs C)
2.37	n.s.	<.001***	<.001***

*** p<.001

n.s. indicate non-significant findings

Appendix R: About the Authors

CHEW Han Ei leads the Governance & Economy cluster at the Institute of Policy Studies (IPS). His work focuses on quantitative policy research, with a strong interest in online harms, digital trust and technology adoption.

He has served as Principal Investigator for multiple large-scale research grants and collaborates closely with public agencies to inform decision-making and policy development. His approach is empirically grounded and hands-on — from designing social science research projects to leading data analyses that shape real-world outcomes.

Outside of IPS, Han Ei serves on the board of SG Her Empowerment and is a pro bono Research Consultant to UNESCO. Some of his key international projects for UNESCO include “Reading in the Mobile Era: A Study of Mobile Reading in Developing Countries” and “I’d Blush If I Could: Closing Gender Divides in Digital Skills through Education”.

Han Ei earned his PhD in Media and Information Studies from Michigan State University. He also writes *The Chart Doctor Has Issues* — a Substack newsletter on data storytelling, visual best practices and the occasional chart takedown.

Carol **SOON** is Deputy Head and Associate Professor (Practice) at the Department of Communications and New Media in the National University of

Singapore (NUS). She is a member of the World Economic Forum's Global Future Council on Information Integrity (2025 to 2026) and Principal Investigator at the NUS Centre for Trusted Internet and Community. She is also Adjunct Principal Scientist at the Centre for Advanced Technologies in Online Safety (CATOS) that is set up by Singapore's Ministry of Digital Development and Information.

Her research interests are in media regulation and digital policy, social media governance, digital literacy and policy communication. Dr Soon has written over 50 media commentaries on the impact of technology, media regulation, digital literacy and digital upskilling of citizens and workers, and public communication. She has single-authored and co-authored more than 80 research reports, journal articles, book chapters and conference papers. Her book, *Mobile Communication and Online Falsehoods: Trends, Impact and Practice*, published by Springer Nature in 2023, addresses existing gaps in research and practice in the management of online falsehoods on instant messaging platforms in Asia.

Dr Soon is currently Vice Chair of Singapore's Media Literacy Council. She is also a member of the Ministry of Culture, Community and Youth's Co-Governance Community of Practice, and serves on the National Crime Prevention Council.

Harkiran **KAUR** is a research assistant at the Institute of Policy Studies (Lee Kuan Yew School of Public Policy, National University of Singapore). Her

research focuses on the social and policy impacts of digital media and the internet, covering topics like digital literacy, online safety and artificial intelligence. She holds a Bachelor's degree in Social Sciences from Singapore Management University.

About IPS Working Paper Series

The IPS Working Papers Series is published in-house for early dissemination of works-in-progress. This may be research carried out by IPS researchers, work commissioned by the Institute or work submitted to the Institute for publication.

The views expressed in the Working Papers are strictly those of the author(s) alone and do not necessarily reflect the views of the IPS.

Comments on the Working Papers are invited. Please direct your comments and queries to the author(s).

IPS Working Papers are available from the IPS at $7.00 each (before GST). Postage and handling charges will be added for mail orders.

For more information, please visit www.lkyspp.nus.edu.sg/ips or
contact email: ips@nus.edu.sg.

Institute of Policy Studies
Lee Kuan Yew School of Public Policy National University of Singapore
1C Cluny Road House 5
Singapore 259599

Web: www.lkyspp.nus.edu.sg/ips
Registration Number: 200604346E