PARLIAMENT OF THE COMMONWEALTH OF AUSTRALIA

Social Media and Online Safety

House of Representatives Select Committee on Social Media and Online Safety

March 2022
CANBERRA

© Commonwealth of Australia

ISBN 978-1-76092-373-0 (Printed Version)

ISBN 978-1-76092-374-7 (HTML Version)

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Australia License.

The details of this licence are available on the Creative Commons website: http://creativecommons.org/licenses/by-nc-nd/3.0/au/.

Contents

Chair's Foreword

Many stories simply just break your heart.

Tilly's story, especially, broke mine. A young teenage girl from regional NSW with everything to live for, she was kind, vibrant and gentle. Tilly loved to dance and paint. She was a star debater.

Tragically, Tilly Rosewarne passed after suffering horrible online harm that no person should ever have to endure.

Even with all the troubling stories the Committee heard through extensive public hearings and heart felt submissions, Tilly's story was gut wrenching.

In part because she was only 15-years old. But also, because it was so recent: she passed only weeks ago, midway through the Committee's inquiry process.

I spoke with her mother Emma about Tilly's experience, and while not forming part of the evidence gathered, Emma asked that her daughter's story be shared.

"We want to do everything we can to make sure no other little humans go through this," Emma told the *Daily Telegraph*.[1]

So this Chair's foreword is dedicated to Tilly, and to every Australian who has suffered harm in the online environment, causing a devastating trail of trauma for victims, their families and their communities.

Tilly's story mirrors that of too many others and is why this inquiry is so important, timely and urgent.

Things must change.

[1] https://www.dailytelegraph.com.au/news/nsw/youth-suicide-death-of-bathurst-schoolgirl-matilda-rosewarne-sparks-call-for-change/news-story/e837015af8a1441b28d87902437e0e3d

The recommendations in this report are an important next step in making our online world and social media platforms safer for all. Too many Australians have been subject to abuse online. Stories of harm, intimidation, and trolling are simply not acceptable.

In the words of Tilly's parents, "Every post you write, every image you share, every word you say, has an impact. We beg you, before you post, share or speak, ask yourself three questions: is it true? Is it kind? Is it necessary? If the answer is "no" to any of these questions, do not post, do not share or do not speak. Because no one ever know when these actions are that human's deepest cut… or the last cut."[2]

During public hearings, the Committee heard first-hand how online abuse can impact all Australians, regardless of age, gender or background. Powerful evidence given by victims of online abuse demonstrated that such behaviour has the capacity to cause significant and long-term harm. Vulnerable users, such as children and young people, women, and those from culturally diverse backgrounds, are at heightened risk of abuse and are more likely to suffer effects from harm such as impacts to mental health and trauma.

Social media companies have made efforts to address these issues through changes to their platforms, including amendments to their terms of use, rules and standards. The Australian Government has also passed world-first legislation which empowers the eSafety Commissioner to tackle online abuse, and new legislative initiatives have also provided the eSafety Commissioner with heightened powers to impose standards and community expectations on platforms.

Notwithstanding these developments, more must be done to ensure the safety of all Australians in online environments.

For too long social media platforms have been able to 'set the rules', enabling the proliferation of online abuse on their spaces. The balance of responsibility for the safety of users online, which until recently has been primarily on users, must be 'flipped' to ensure that social media platforms bear the ultimate burden of providing safety for their users.

The inquiry's findings fall into three categories: what industry can do, what government can do, and the 'missing middle' – what individuals and society at large can do to address online harm. The latter category particularly addresses

[2] Quote provided by the Rosewarne Family and as seen in https://www.dailytelegraph.com.au/news/nsw/youth-suicide-death-of-bathurst-schoolgirl-matilda-rosewarne-sparks-call-for-change/news-story/e837015af8a1441b28d87902437e0e3d

online culture which allows harm to proliferate, and how we can prevent harm from happening in the first place.

It became apparent throughout the inquiry that while technology, social media platforms and government have a role to play in addressing online harm, there is also a need to focus on the conduct and behaviour of individuals who use technology in ways that harm others.

No technology can prevent a decision to deliberately bully another person online. We cannot legislate people's intentions towards others, but we can look at the behaviours of Australians themselves, the responsibility we all have to be respectful online and the consequences for engaging and perpetrating harm to others.

The Committee is cognisant of addressing the issues faced by victims of online abuse and is also aware that there cannot be a one-size-fits-all approach, and that this work is ongoing as technology and the ways humans use it evolves. Hence, this topic deserves further consideration over a much longer timeframe.

I want to thank witnesses who provided evidence to the Committee, particularly those who shared powerful personal accounts of the impact of online abuse. I thank my parliamentary colleagues from the Joint Committee on the Draft Online Safety Bill in the United Kingdom Parliament who provided the Committee with insight into international regulatory responses.

This inquiry has been extensive, involving 11 public hearings and a number of private meetings. The issues explored were both broad and important to the lives of everyday Australians. All Committee members, and in particular the Deputy Chair, Tim Watts MP, have given generously of their time while also approaching this inquiry with an open mind and a genuine desire to pursue change to create a safer online environment.

The Secretariat team has been outstanding in their professionalism, dedication and genuine commitment throughout this inquiry. The Committee gathered a range and depth of information that would usually be the purview of a much longer inquiry. The Secretariat's expertise has assisted Committee members beyond measure.

Finally, I thank every individual who showed courage, vulnerability and bravery in speaking up and speaking out. It is the poignancy of their insights based on personal experience that has guided the recommendations in this report aimed at ensuring the online world is a safer place for all.

Ms Lucy Wicks MP

Member for Robertson

Members

Chair

Mrs Lucy Wicks MP	Robertson, NSW

Deputy Chair

Mr Tim Watts MP	Gellibrand, VIC

Members

Ms Sharon Claydon MP	Newcastle, NSW
Ms Nicolle Flint MP (until 21.12.21)	Boothby, SA
Ms Celia Hammond MP (from 21.12.21)	Curtin, WA
Mr Craig Kelly MP	Hughes, NSW
Mr Dave Sharma MP	Wentworth, NSW
Mr Julian Simmonds MP	Ryan, QLD
Dr Anne Webster MP	Mallee, VIC

Committee Secretariat

Joel Bateman, Committee Secretary

Ophelia Tynan, A/g Inquiry Secretary

Peter Pullen, Senior Researcher

Mihira Jeyanarayanan, Researcher

Carissa Skinner, Office Manager

Terms of Reference

The Select Committee on Social Media and Online Safety will inquire into:

a) the range of online harms that may be faced by Australians on social media and other online platforms, including harmful content or harmful conduct;

b) evidence of:

   i) the potential impacts of online harms on the mental health and wellbeing of Australians;
   ii) the extent to which algorithms used by social media platforms permit, increase or reduce online harms to Australians;
   iii) existing identity verification and age assurance policies and practices and the extent to which they are being enforced;

c) the effectiveness, take-up and impact of industry measures, including safety features, controls, protections and settings, to keep Australians, particularly children, safe online;

d) the effectiveness and impact of industry measures to give parents the tools they need to make meaningful decisions to keep their children safe online;

e) the transparency and accountability required of social media platforms and online technology companies regarding online harms experienced by their Australians users;

f) the collection and use of relevant data by industry in a safe, private and secure manner;

g) actions being pursued by the Government to keep Australians safe online; and

h) any other related matter.

Abbreviations

ACCCE	Australian Centre to Counter Child Exploitation
ACMA	Australian Communications and Media Authority
AEC	Australian Electoral Commission
AFL	Australian Football League
AFLW	Australian Football League Women's
AFP	Australian Federal Police
AGD	Attorney-General's Department
AHRC	Australian Human Rights Commissioner
AI	Artificial intelligence
AMF	Alannah and Madeline Foundation
ANROWS	Australia's National Research Organisation for Women's Safety
APP	Australian Privacy Principles
BOSE	Basic Online Safety Expectations
CALD	Culturally and linguistically diverse
CCSP	Council of Catholic School Parents NSW ACT
CDW	Centre for Digital Wellbeing
CEO	Chief Executive Officer
COVID-19	Novel coronavirus disease (SARS-CoV-2)
CRF	Carly Ryan Foundation
CSAM	Child sexual abuse material

DIGI	Digital Industry Group Inc.
DMF	Daniel Morcombe Foundation
DRW	Digital Rights Watch
E2EE	End-to-end encryption
EOS Act	*Enhancing Online Safety Act 2015* (Cth)
eSafety	Office of the eSafety Commissioner
EU	European Union
Giggle	Giggle for Girls Pty Ltd
HDC Act	*Harmful Digital Communications Act 2015* (New Zealand)
HDC Bill	Harmful Digital Communications (Unauthorised Posting of Intimate Visual Recording) Amendment Bill (New Zealand)
Home Affairs	Department of Home Affairs
Infrastructure	Department of Infrastructure, Transport, Regional Development and Communications
IPO Act	*Telecommunications Legislation Amendment (International Production Orders) Act 2021*
JCDOSB	Joint Committee on the Draft Online Safety Bill (United Kingdom)
LGBTIQA+	Lesbian, Gay, Bisexual, Transgender, Intersexual, Queer, Asexuality and Other
Meta	Meta Platforms
NMHC	National Mental Health Commission
NSCPCC	National Society for the Prevention of Cruelty to Children
OSA	*Online Safety Act 2021* (Cth)
Privacy Act	*Privacy Act 1988*
RDA	*Racial Discrimination Act 1975* (Cth)
SLAID Act	*Surveillance Legislation Amendment (Identify and Disrupt) Act 2021*
Snap	Snap Inc.
UK	United Kingdom
US	United States of America
WESNET	Women's Services Network

List of Recommendations

Recommendation 1

1.25 The Committee recommends that the Australian Government propose the appointment of a House Standing Committee on Internet, Online Safety and Technological Matters, from the commencement of the next parliamentary term.

Recommendation 2

1.26 The Committee recommends that, subject to Recommendation 1, the Australian Government propose an inquiry into the role of social media in relation to democratic health and social cohesion, to be referred to the aforementioned committee or a related parliamentary committee.

Recommendation 3

2.145 The Committee recommends that the eSafety Commissioner undertakes research focusing on how broader cultural change can be achieved in online settings.

Recommendation 4

2.146 Subject to the findings in Recommendation 3, the Committee recommends that the Australian Government establishes an educational and awareness campaign targeted at all Australians, focusing on digital citizenship, civics and respectful online interaction.

Recommendation 5

2.147 The Committee recommends that the eSafety Commissioner examine the extent to which social media companies actively prevent:

- recidivism of bad actors,
- pile-ons or volumetric attacks, and
- harms across multiple platforms.

2.148 The eSafety Commissioner should then provide the Australian Government with options for a regulatory framework, including penalties for repeated failures.

Recommendation 6

2.154 The Committee recommends that the Office of the eSafety Commissioner be provided with adequate appropriations to establish and manage an online single point of entry service for victims of online abuse to report complaints and be directed to the most appropriate reporting venue, dependent on whether their complaints meet the requisite threshold, and in consideration of a variety of audiences such as children, parents/carers, women, people from culturally and linguistically diverse backgrounds, and other relevant vulnerable groups.

Recommendation 7

2.160 The Committee recommends that the Australian Government refer to the proposed House Standing Committee on Internet, Online Safety and Technological Matters, or another committee with relevant focus and expertise, an inquiry into technology-facilitated abuse, with terms of reference including:

- The nature and prevalence of technology-facilitated abuse;
- Responses from digital platforms and online entities in addressing technology-facilitated abuse, including how platforms can increase the safety of their users; and
- How technology-facilitated abuse is regulated at law, including potential models for reform.

Recommendation 8

2.161 The Committee recommends that the Australian Government significantly increase funding to support victims of technology-facilitated abuse, through existing Australian Government-funded programs. This should include additional funding for specialised counselling and support services for victims; and be incorporated in the next National Action Plan to End Violence Against Women and Children 2022-2032.

Recommendation 9

3.185 The Committee recommends that future reviews of the operation of the *Online Safety Act 2021* take into consideration the implementation of the Safety by Design Principles on major digital platforms, including social media services and long-standing platforms which require retrospective application of the Safety by Design Principles.

Recommendation 10

3.197 The Committee recommends that the Department of Infrastructure, Transport, Regional Development and Communications, in conjunction with the eSafety Commissioner and the Department of Home Affairs, examine the need for potential regulation of end-to-end encryption technology in the context of harm prevention.

Recommendation 11

3.198 The Committee recommends that the eSafety Commissioner, as part of the drafting of new industry codes and implementation of the Basic Online Safety Expectations:

- Examine the extent to which social media services adequately enforce their terms of service and community standards policies, including the efficacy and adequacy of actions against users who breach terms of service or community standards policies;

- Examine the potential of implementing a requirement for social media services to effectively enforce their terms of service and community standards policies (including clear penalties or repercussions for breaches) as part of legislative frameworks governing social media platforms, with penalties for non-compliance; and

- Examine whether volumetric attacks may be mitigated by requiring social media platforms to maintain policies that prevent this type of abuse and that require platforms to report to the eSafety Commissioner on their operation.

Recommendation 12

3.199 The Committee recommends that the eSafety Commissioner examine the extent to which social media companies actively apply different standards to victims of abuse depending on whether the victim is a public figure or requires a social media presence in the course of their employment, and provides options for a regulatory solution that could include additions to the Basic Online Safety Expectations.

Recommendation 13

3.204 The Committee recommends that the eSafety Commissioner, in conjunction with the Department of Infrastructure, Transport, Regional Development and Communications and the Department of Home Affairs and other technical experts as necessary, conduct a review of the use of algorithms in digital platforms, examining:

- How algorithms operate on a variety of digital platforms and services;
- The types of harm and scale of harm that can be caused as a result of algorithm use;
- The transparency levels of platforms' content algorithms;
- The form in which regulation should take (if any); and
- A roadmap for Australian Government entities to build skills, expertise and methods for the next generation of technological regulation in order to develop a blueprint for the regulation of Artificial Intelligence and algorithms in relation to user and online safety, including an assessment of current capacities and resources.

Recommendation 14

3.205 The Committee recommends that the eSafety Commissioner require social media and other digital platforms to report on the use of algorithms,

detailing evidence of harm reduction tools and techniques to address online harm caused by algorithms. This could be achieved through the mechanisms provided by the Basic Online Safety Expectations framework and Safety By Design assessment tools, with the report being provided to the Australian Government to assist with further public policy formulation.

Recommendation 15

3.213 The Committee recommends that, subject to Recommendation 19, the proposed Digital Safety Review make recommendations to the Australian Government on potential proposals for mandating platform transparency.

Recommendation 16

3.218 The Committee recommends the implementation of a mandatory requirement for all digital services with a social networking component to set default privacy and safety settings at their highest form for all users under 18 (eighteen) years of age.

Recommendation 17

3.219 The Committee recommends the implementation of a mandatory requirement for all technology manufacturers and providers to ensure all digital devices sold contain optional parental control functionalities.

Recommendation 18

4.110 The Committee recommends that the Department of Infrastructure, Transport, Regional Development and Communications conduct a Digital Safety Review on the legislative framework and regulation in relation to the digital industry. The Digital Safety Review should commence no later than 18 months after the commencement of the *Online Safety Act 2021*, and provide its findings to Parliament within twelve (12) months.

Recommendation 19

4.111 The Committee recommends that, subject to Recommendation 18, the Digital Review examine the need and possible models for a single regulatory framework under the Online Safety Act, to simplify regulatory arrangements.

Recommendation 20

5.88 The Committee recommends that the Digital Review include in its terms of reference:

- The need to strengthen the Basic Online Safety Expectations to incorporate and formalise a statutory duty of care towards users;
- The scope and nature of such a duty of care framework, including potential models of implementation and operation;
- Potential methods of enforcement to ensure compliance, including penalties for non-compliance; and
- The incorporation of the best interests of the child principle as an enforceable obligation on social media and other digital platforms, including potential reporting mechanisms.

Recommendation 21

5.92 The Committee recommends that the eSafety Commissioner:

- Increase the reach of educational programs geared at young people regarding online harms, with a particular focus on reporting mechanisms and the nature of some online harms being a criminal offence;
- Formalise a consultation and engagement model with young people through the Australian Government's Youth Advisory Council in regards to educational themes and program delivery; and
- Report to the Parliament on the operation and outcomes of the program, including research identifying whether this has resulted in a reduction in online harm for young people.

Recommendation 22

5.99 The Committee recommends that the eSafety Commissioner work in consultation with the Department of Education, Skills and Employment to design and implement a national strategy on online safety education

designed for early childhood, and primary school-aged children, and secondary school-aged young people, including:

- A proposed curriculum, informed by developmental stages and other relevant factors;
- Potential methods of rollout, including consultation and engagement with children, young people, child development and psychology experts, digital education experts and other specialists in online harm; and
- A roadmap provided to parents of these age groups detailing methods of addressing online harm.

Recommendation 23

5.100 The Committee recommends that the eSafety Commissioner design and administer an education and awareness campaign aimed at adults, particularly in relation to vulnerable groups such as women, migrant and refugee groups, and people with disabilities, with a focus on the eSafety Commissioner's powers to remove harmful content and the mechanisms through which people can report harmful content and online abuse.

Recommendation 24

5.101 The Committee recommends that the Australian Government work with states and territories to ensure that relevant law enforcement agencies are appropriately trained on how to support victims of online harm. This should include trauma-informed approaches as well as a comprehensive understanding of police powers and other relevant avenues, such as the relevant powers of the eSafety Commissioner.

Recommendation 25

5.102 The Committee recommends that the Australian Government review funding to the eSafety Commissioner within twelve (12) months to ensure that any of the Committee's recommendations that are agreed to by the Government and implemented by the Office of the eSafety Commissioner are adequately and appropriately funded for any increased resource requirements.

Recommendation 26

5.103 The Committee recommends that the Online Safety Youth Advisory Council, via the eSafety Commissioner, provide a response to this report and its recommendations within six (6) months of its establishment and full membership.

1. Introduction

Details of the Inquiry

1.1 The impact of the online world in Australians' lives cannot be overestimated. According to the Australian Bureau of Statistics, in 2016-17, approximately 86 per cent of Australian households had internet access, and 97 per cent of households with children under the age of 15 years of age.[1] Australians engage with friends, family, businesses and government online, in ever-increasing numbers and ways.

1.2 Notwithstanding the positives that online services and products can provide, there are also significant and serious dangers present in digital spaces, at rates never seen before. Online safety has become an international issue, both in the transnational way corporations and online actors engage in the digital world, and for governments globally in addressing these matters.

1.3 The regulation of online spaces has rapidly evolved in a short space of time. A number of frameworks have been established within the past five to ten years to moderate the digital industry. Despite this, concerns remain that more can, and should, be done to keep Australians safe online, particularly the most vulnerable of internet users.

1.4 This inquiry considered social media and online safety for Australians engaging in the digital space. It examined the harms that Australians face, the nature of the online world where they experience them and the actors

[1] Australian Bureau of Statistics, *Household Use of Information Technology*, Australia, 2016-17, available at: https://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0 (accessed 2 February 2022).

who participate in it, and the current legislative framework and government policies that seek to regulate online safety. The inquiry also examined potential ways forward to improve online safety, particularly for vulnerable users.

1.5 This chapter provides an overview of the inquiry's scope and conduct. It also discusses options for potential avenues for future parliamentary scrutiny of digital matters.

Recent inquiries

1.6 As evidence of the complexity and evolution of these issues, multiple parliamentary inquiries and reviews have been conducted in relation to online safety matters within the past decade. Table 1.1 lists relevant parliamentary inquiries which considered online safety issues from 2015 to 2022.

1.7 These inquiries have canvassed a broad range of topics, including specific types of harm (e.g. pornography, extremism, child exploitation), technological tools to prevent or minimise harm (e.g. age verification technology), and potential law reform to better address emerging online threats (e.g. the adequacy of current laws in preventing cyberbullying and image-based abuse).

Table 1.1 Parliamentary inquiries considering online safety from 2015 to 2022[2]

Year of tabling	Committee	Inquiry
Current	Joint Standing Committee on Intelligence and Security	Inquiry into extremist movements and radicalism in Australia
	Joint Committee on Law Enforcement	Law enforcement capabilities in relation to child exploitation
	---	Impact of illicit drugs being traded online
2020	House Standing Committee on Social Policy and Legal Affairs	Inquiry into age verification for online wagering and online pornography

[2] This table excludes inquiries into bills presented to Parliament and inquiries without a strong relation to online harm.

	Senate Select Committee on Foreign Interference through Social Media	The risk posed to Australia's democracy by foreign interference through social media
	Joint Committee on Law Enforcement	The impact of new and emerging information and communications technology on Australian law enforcement agencies
2018	Senate Standing Committee on Legal and Constitutional Affairs (References)	Adequacy of existing offences in the Commonwealth Criminal Code and of state and territory criminal laws to capture cyberbullying
2017	Senate Standing Committee on Environment and Communications (References)	Participation of Australians in online poker
2016	---	Harm being done to Australian children through access to pornography on the Internet
	Senate Standing Committee on Legal and Constitutional Affairs (References)	Phenomenon colloquially referred to as 'revenge porn', which involves sharing private sexual images and recordings of a person without their consent, with the intention to cause that person harm
2015	House Standing Committee on Communication and the Arts	Inquiry into the use of subsection 313(3) of the Telecommunications Act 1997 by government agencies to disrupt the operation of illegal online services

Source: Australian Parliament House website.

1.8 In addition, on 10 February 2022 the Senate referred the provisions of the Social Media (Anti-Trolling) Bill 2022 to the Senate Standing Committee on

Legal and Constitutional Affairs. The inquiry into the bill's provisions is required to be tabled in the Senate by 22 March 2022.[3]

Other government-led reviews

1.9 A number of government-led reviews or consultations relating to aspects of online safety have been conducted or are currently also underway. A non-exhaustive selection of these include:

- Report of the Statutory Review of the Enhancing Online Safety Act 2015 and the Review of Schedules 5 and 7 to the Broadcasting Services Act 1992 (Online Content Scheme), conducted by an independent reviewer and published in October 2018;[4]
- The Digital Platforms Inquiry conducted by the Australian Competition and Consumer Commission, released in June 2019;[5]
- Amendments to the *Privacy Act 1988*;[6] and
- Consultations led by the Attorney-General's Department in relation to the Social Media (Anti-Trolling) Bill.[7]

Scope of the inquiry

1.10 On 1 December 2021, the House of Representatives resolved to establish a Select Committee on Social Media and Online Safety (the Committee) to inquire into matters associated with social media and online safety.

---

3 Senate Standing Committee on Legal and Constitutional Affairs, *About this inquiry – Social Media [Anti-Trolling] Bill 2022*, available at: https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Legal_and_Constitutional_Affairs/Anti-Trolling (accessed 17 February 2022).

4 Lynette Briggs AO, *Report of the Statutory Review of the Enhancing Online Safety Act 2015 and the Review of Schedules 5 and 7 to the Broadcasting Services Act 1992 (Online Content Scheme)*, October 2018, available at: https://www.infrastructure.gov.au/sites/default/files/briggs-report-stat-review-enhancing-online-safety-act2015.pdf (accessed 22 December 2021).

5 Australian Competition and Consumer Commission, *Digital Platforms Inquiry – Final Report*, June 2019.

6 Senator the Hon Michaelia Cash, Attorney-General, and the Hon David Coleman MP, Assistant Minister to the Prime Minister for Mental Health and Suicide Prevention, 'Landmark privacy reforms to better protect Australians online', joint media release, 25 October 2021, available at: https://ministers.ag.gov.au/media-centre/landmark-privacy-reforms-better-protect-australians-online-25-10-2021 (accessed 23 February 2022).

7 Attorney-General's Department, *Social Media (Anti-Trolling) Bill*, available at: https://www.ag.gov.au/legal-system/social-media-anti-trolling-bill (accessed 4 January 2022).

1.11 The terms of reference, which can be found at page xxi of this report, enabled the Committee to take a broad approach in examining matters pertaining to online safety and the role of social media. The Committee heard evidence in relation to a range of topics involving online safety, including risks to vulnerable users such as children, women and minority groups, legislative frameworks and current government responses to address aspects of online safety.

Definition of 'online safety' and 'social media'

1.12 The inquiry's terms of reference refer to both 'online safety' and 'social media'. This section outlines definitions for these terms, and the approach taken by the Committee when using these definitions to conduct the inquiry.

'Online safety'

1.13 Chapter 2 outlines the legislative definition of 'online safety' as determined by the *Online Safety Act 2021* (Cth) (the OSA) in addition to legislative definitions for cyberbullying. The OSA defines 'online safety for Australians' as 'the capacity of Australians to use social media services and electronic services in a safe manner'.[8]

1.14 This definition provides an outline of what constitutes online safety, as opposed to what is not safe in an online environment. 'Online harm', in contrast, is an extremely broad category and captures a wide range of online behaviour in multiple forms and platforms.

1.15 Due to the short timeframe in which the Committee was required to conduct its inquiry and report its findings, the Committee was unable to examine many other critical issues relating to online safety. Further, the Committee did not wish to 'reinvent the wheel' by examining topics which have been thoroughly canvassed in other inquiries by parliamentary committees in greater depth.

1.16 Topics that were not examined in depth include: the use of social media to harm democracy and social cohesion, disinformation and misinformation in online platforms and social media (particularly in relation to medical information), foreign interference and state actors on social media, the role

[8] *Online Safety Act 2021* (Cth), s 5. A second definition is included specifically in relation to online safety for children in section 6, which replicates the definition in section 5 but adds 'and includes the protection of Australian children using those services from cyber-bullying material targeted at an Australian child'.

of the dark web in online safety, and issues in relation to political communication online.

1.17 Further, online safety is not applicable only to social media platforms. Internet users are exposed to an ever-increasing variety of harms online, some of which are not exclusive to social media services. In addition, while this Committee has examined the actions of the 'big tech' social media companies, it noted witnesses' comments in relation to the responsibilities and actions of other forms of technology providers and platforms, as well as smaller social media companies.

'Social media'

1.18 Social media is defined in the OSA as:

> (a) an electronic service that satisfies the following conditions:
>
> i. the sole or primary purpose of the service is to enable online social interaction between 2 or more end-users;
>
> ii. the service allows end-users to link to, or interact with, some or all of the other end-users;
>
> iii. the service allows end-users to post material on the service;
>
> iv. such other conditions (if any) as are set out in the legislative rules; or
>
> (b) an electronic service specified in the legislative rules[9].

1.19 This definition is inclusive of an enormous range of current and emerging services online. An inquiry into social media and online safety cannot exist in a vacuum without consideration of other forms of online platforms, such as search engines, news media sites, e-commerce marketplaces and so on. Further, digital platforms as they currently exist often defy categorisation into a single type, meaning that many platforms not advertised as social media services can incorporate elements and technology similar to that used by the major social media firms.

1.20 While the Committee was mindful of the changing and complex nature of services that exist online, it did not focus specific attention on other arenas of digital life.

---

9 *Online Safety Act 2021*, section 13(1).

Parliamentary oversight of online matters

1.21 Issues in relation to online safety have been of interest to parliamentary committees since the internet's rise in modern society. Despite this, the topic does not fit neatly within the oversight of one particular committee.

1.22 Matters in relation to internet security and safety are within the purview of both the Standing Committee on Communications and the Arts and the Standing Committee on Industry, Innovation, Science and Resources. Both Committees have conducted inquiries in relation to internet commerce and technology in the past. Further, the Joint Committee on Law Enforcement also regularly conducts inquiries on online-related matters, such as its review of the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*[10] as does the House Standing Committee on Social Policy and Legal Affairs, which recently conducted an inquiry into age verification for online wagering and online pornography.[11]

1.23 Parliamentary interest in relation to the internet industry is unlikely to wane as Australians' lives progressively transfer online. Accordingly, the Committee considers it appropriate that a Standing Committee on internet and e-commerce matters be established to address the uncertain space where these matters lie. A dedicated committee could attract members with specialist knowledge or interest in relation to technological development and innovation. It would also ensure that the Parliament is informed of the latest developments, risks, and solutions relevant to the industry.

1.24 The Committee believes that further examination of other issues should be addressed by this new committee, including:

- A follow-up inquiry to examine the responses of interested parties to the recommendations contained in this report;
- A review of the rollout of the OSA, including responses from industry and the community;

---

10 Joint Committee on Law Enforcement, *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*, December 2021, available at: https://www.aph.gov.au/Parliamentary_Business/Committees/Joint/Law_Enforcement/AVMAct/Report (accessed 31 January 2022).

11 House Standing Committee on Social Policy and Legal Affairs, *Protecting the age of innocence*, February 2020, available at: https://www.aph.gov.au/Parliamentary_Business/Committees/House/Social_Policy_and_Legal_Affairs/Onlineageverification/Report (accessed 31 January 2022).

- The operation and role of algorithms in social media and other digital services, including potential methods of regulation and positive uses of algorithms to improve online safety;
- The role of social media in relation to democratic health and social cohesion;
- Cryptocurrency, its use in illegal activities online, and potential regulatory activities, including the use of non-fungible tokens; and
- The role of online gaming platforms in online safety, including cyberbullying and child sexual abuse material.

Recommendation 1

1.25 The Committee recommends that the Australian Government propose the appointment of a House Standing Committee on Internet, Online Safety and Technological Matters, from the commencement of the next parliamentary term.

Recommendation 2

1.26 The Committee recommends that, subject to Recommendation 1, the Australian Government propose an inquiry into the role of social media in relation to democratic health and social cohesion, to be referred to the aforementioned committee or a related parliamentary committee.

Conduct of the inquiry

1.27 A media release announcing the inquiry was issued on 13 December 2021, calling for submissions from interested individuals and organisations on the terms of reference. The Committee also directly invited submissions from industry bodies, agencies, institutions, academics, think tanks and individuals.

1.28 On 15 February 2022, the House of Representatives agreed to an extension of time for the Committee to provide its report, amending the final reporting date to 15 March 2022.

1.29 The inquiry received 107 submissions (including supplementary submissions) and 12 exhibits, which are listed at Appendix A and B respectively.

1.30 The Committee held eleven public hearings, hearing from 58 separate witnesses (consisting of individuals and organisations) during the course of the inquiry. A list of those hearings and the witnesses and organisations that

appeared at the hearings are listed at Appendix C. All hearings were held via videoconference and/or teleconference from Canberra. The Committee was also able to undertake two virtual site visits, as well as meet with members of a delegation from the European Parliament.

1.31 The Committee notes that some of the submissions received as part of this inquiry, as well as some of the hearing transcripts, contain references to or examples of a range of harms experienced online, and advises members of the public to bear that in mind when reviewing the inquiry's evidence.

1.32 The Committee thanks all those who participated in the inquiry for giving their time to provide evidence, either through a written submission or at a public hearing. The Committee especially thanks the witnesses who participated in hearings from locations around the world, often at unsociable hours, and is grateful for their valuable testimony.

1.33 Finally, the Committee particularly acknowledges the bravery demonstrated by individuals who provided the Committee with their personal experiences in relation to online harm. These experiences demonstrated to the Committee the seriousness and magnitude of the harm that online abuse and other forms of harm can cause to Australians. The Committee thanks these individuals for their powerful stories which have informed the findings contained in this report.

2. You Have A New Notification

The Forms and Impacts of Online Harm

Overview

2.1 The short time in which the internet has been a presence in the lives of everyday Australians has been transformative in unprecedented and extraordinary ways. One witness described the impact of digitisation as 'akin to what happened in the industrial revolution centuries ago. It, effectively, is changing anything and everything that we do'.[1]

2.2 Digital services offer Australians a vast number of benefits, including communication, work and educational opportunities, networking, e-commerce and others. Accordingly, in examining online harm, the Committee is cautious to avoid making a blanket assumption or finding that the entirety of the online world, or particular online platforms, are inherently negative or damaging.

2.3 Nonetheless, the Committee heard extensive evidence suggesting that online harm is rampant on digital spaces. Victims of online abuse indicated that harmful content in digital settings had resulted in significant and lasting impacts, ranging from psychological harm to the impact on life choices such

[1] Ms Christine Morgan, Chief Executive Officer and Prime Minister's National Suicide Prevention Adviser, National Mental Health Commission (NMHC), *Committee Hansard*, 21 January 2022, p. 8.

as careers to fears for personal safety. This situation constitutes an urgent threat to the digital and personal wellbeing of Australians.

2.4 This chapter examines the nature of online harm. It then outlines who is most at risk of being a victim of online harm, and examines vulnerable groups that are at particular risk for online harm. The chapter concludes by examining the effects of online harm on victims.

2.5 *Readers are advised that this chapter contains material that may be distressing, including references to and examples of forms of abuse.*

The digital world as a medium for mixed experiences

2.6 In examining online harm, it is important to recognise that not all interactions in the digital space are negative or abusive. For many Australians, the online world has been a source of positivity and social connection.

2.7 The Office of the eSafety Commissioner (eSafety) strongly advocated this message in its submission, stating that the internet and social media platforms offer users with countless sources of positive benefits:

> Social media connects people with the world around them, as well as with their communities and their families. The stark and isolating nature of the pandemic has crystalised the need to access these channels. Forty-nine percent of Australians were either born overseas or have families overseas, and the online world can help keep them connected. Similarly, online connectivity remains critical for regional and remote communities, and for our Aboriginal and Torres Strait Islander communities who want to remain connected to country and culture. In addition, we know the benefits to neurodiverse young people from engaging online are evident both inside and outside of formal education. Being online helps them to develop social skills and offers ways to expand and enrich offline interests.[2]

2.8 The National Mental Health Commission (the NMHC) highlighted that digital platforms can be a source of good experiences as well as harmful ones, and that society needed to move towards a form of engagement with social media in particular which 'enhances mental health and wellbeing'.[3] The NMHC pointed out many benefits of online engagement, such as social engagement, education and employment opportunities.[4] Further, studies

---

2 eSafety Commissioner, *Submission 53*, p. 14.

3 Ms Christine Morgan, NMHC, *Committee Hansard*, 21 January 2022, p. 5.

4 Ms Christine Morgan, NMHC, *Committee Hansard*, 21 January 2022, pp 6-7.

indicate that children and young people predominantly use social media for 'communication, connection and sharing with others'.[5] The NMHC confirmed this, stating that community and social engagement was of critical importance to individual and societal wellbeing, as was seen particularly clearly during the COVID-19 pandemic.[6]

2.9 The Alannah and Madeline Foundation (AMF) agreed with this statement, stating that children and young people in particular use internet services and products to explore their personal development.[7] The NMHC similarly reported young people's experiences of finding their 'tribe' online, which they argued was particularly important due to the widely dispersed geography of Australia, and the consequent social isolation for those in rural and regional communities.[8]

2.10 While online spaces have the potential for positive experiences, a broad range of negative experiences were reported to the Committee. The remainder of this chapter examines online harm in depth.

Definition of 'online safety' and 'online harm'

2.11 An internationally accepted definition of 'online harm', including types and definitions of common forms of harm, currently does not exist.[9] The eSafety Commissioner noted that the *Online Safety Act 2021* (Cth) (the OSA) sets out specific types of online harm, but that:

> At this stage, it is largely up to individual online service providers to establish rules and guidelines for this type of activity and content that is or is not permitted on their platforms within community guidelines or terms of service. However, these can diverge significantly across services.[10]

2.12 As stated in Chapter 1, the OSA defines 'online safety for Australians' as 'the capacity of Australians to use social media services and electronic services in

---

5 Professor Amanda Third, Professorial Research Fellow, Institute for Culture and Society, Western Sydney University; Co-Director, Young and Resilient Research Centre, Western Sydney University (Young and Resilient Centre), *Committee Hansard*, 21 December 2021, p. 26.

6 Ms Christine Morgan, NMHC, *Committee Hansard*, 21 January 2022, p. 6.

7 Ms Sarah Davies, Chief Executive Officer, Alannah and Madeline Foundation (AMF), *Committee Hansard*, 21 December 2021, p. 23.

8 Ms Christine Morgan, NMHC, *Committee Hansard*, 21 January 2022, p. 7.

9 eSafety Commissioner, *Submission 53*, p. 26.

10 eSafety Commissioner, *Submission 53*, p. 26.

a safe manner'.[11] 'Online harm' can therefore encompass a broad range of conduct across differing platforms and online spaces, which is ever-changing in a highly dynamic digital environment.

Typology of online harm

2.13 Interacting with the online world can result in different types of harm. The three main recognised types of harm are:

- Individual harm;
- Community-based harm; and
- Economic harm.

2.14 This chapter will examine these types of harm, focusing in particular on individual harm.

Individual harm

2.15 Individual harm arises due to an individual person's interactions on the internet. Harm can also be experienced while engaging with services such as social media, visiting websites containing illegal or disturbing content, and other situations. A non-exhaustive range of situations that can cause harm include:

- The production and distribution of child sexual abuse material (CSAM), including children and young people being targeted or 'groomed' by perpetrators;
- Cyberbullying, abuse or harassment: Where one person harasses, threatens, intimidates or name-calls another person using an internet service (which also includes volumetric attacks, where a large group of people attack one individual);
- Exposure to illegal or disturbing content, such as child sexual abuse material (CSAM), violent or abhorrent content, or material promoting harmful or dangerous behaviours (e.g. suicide ideation, promotion of eating disorders);
- Discrimination on the basis of sex, gender, sexual orientation, ethnic background, religious belief, political views, and others;

[11] *Online Safety Act 2021* (Cth), s 5. A second definition is included specifically in relation to online safety for children in section 6, which replicates the definition in section 5 but adds 'and includes the protection of Australian children using those services from cyber-bullying material targeted at an Australian child'.

- Technology-facilitated abuse (including the non-consensual distribution of explicit images, deep-fake or cheap-fake image abuse, cyber-flashing, utilising tracking devices or software to monitor a person without consent, and controlling access to accounts or technology); and
- Identity theft or imitation, including people using fake social media accounts of others.

2.16 These forms of harm may not be isolated to a singular type or platform; material can be dispersed across multiple platforms, and victims may be unaware of how far material has circulated or on what platforms material exists.[12]

2.17 The below section provides a brief outline of the main types of individual harm.

Cyberbullying and cyber abuse

2.18 Cyberbullying and cyber abuse[13] includes 'online communication to or about someone which is menacing, harassing or offensive and also intended to cause serious harm to their physical or mental health'.[14]

2.19 The OSA sets out the elements required to establish cyberbullying for adults and children, one form of online harm. For adults, online harassment is described as cyber abuse, which consists of the following elements:

- The material being provided on a social media, relevant electronic, or designated internet service;
- A finding that an ordinary reasonable person would conclude that the material distributed was intended to have an effect of causing serious harm to a particular Australian adult;
- A finding that an ordinary reasonable person in the position of that adult would consider the material being (in all the circumstances) menacing, harassing or offensive; and

---

12 Ms Kara Hinesley, Director of Public Policy, Australia and New Zealand, Twitter, *Committee Hansard*, 21 January 2022, p. 20.

13 The Office of the eSafety Commissioner distinguishes between 'cyberbullying', which is online abuse directed at children, and 'cyber abuse', which is online abuse directed at adults. This distinction demonstrates that the same conduct directed at children may not affect adults in the same way, and vice versa. It also reflects the different ways and arenas that online abuse may be experienced by adults and children.

14 Office of the eSafety Commissioner, *Adult cyber abuse*, available at: https://www.esafety.gov.au/key-issues/adult-cyber-abuse (accessed 28 February 2022).

- Any further conditions set out by the legislative rules.[15]

2.20 For children, similar conduct is described as cyberbullying to reflect how online abuse is experienced differently by children. Section 6 of the OSA outlines the definition of cyberbullying directed at an Australian child, which largely replicates the elements of section 7. The thresholds of what is considered serious harm, however, are lower due to children's developmental stages.

2.21 Both cyber abuse and cyberbullying include a range of behaviours, including:

- abusive texts and emails
- hurtful messages, images or videos
- imitating others online
- excluding others online
- humiliating others online
- spreading nasty online gossip and chat
- creating fake accounts to trick someone or humiliate them.[16]

2.22 Serious forms of online abuse targeting adults can include:

- being harassed and threatened with violence because of their physical appearance, religion, gender, race, disability, sexual orientation or political beliefs
- finding their personal contact details have been made public on a social media service or other online platform in order to scare, harass or attack them
- being threatened with serious harm and other people online being encouraged to join in
- being stalked and threatened online, particularly in the context of domestic and family violence
- being encouraged to harm themselves, particularly in cases where they are known to be at particular risk (for example, because they have a mental health condition)

---

15 *Online Safety Act 2021* (Cth), s 7.

16 Office of the eSafety Commissioner, *Cyberbullying*, available at: https://www.esafety.gov.au/key-issues/cyberbullying (accessed 28 February 2022).

- repeatedly being sent obscene and threatening messages as part of ongoing harassment.[17]

2.23 For cyberbullying targeting children, types of behaviour which fall under the definition used by eSafety includes 'online communication to or about an Australian child which is seriously humiliating, harassing, intimidating or threatening'.[18] Examples of such conduct include 'abusive texts and emails; hurtful messages, images or videos; excluding others; spreading nasty gossip and chat; or creating fake accounts to trick or humiliate someone'.[19]

2.24 Cyberbullying and online abuse can involve large numbers of people or coordinated attacks. eSafety states that this behaviour, known as 'volumetric attacks' (or 'pile-ons' or 'brigades'), can be among the most serious forms of cyberbullying and online abuse.[20]

Technology-facilitated abuse

2.25 Technology-facilitated abuse is a form of domestic and family violence, and is defined as involving 'misuse of devices (such as phones, devices and computers), accounts (such as email) and software or platforms (such as social media) to control, abuse, track and intimidate victim-survivors'.[21]

2.26 Technology-facilitated abuse has now been recognised by the United Nation's Special Rapporteur on violence against women, its causes and consequences as having 'facilitated new types of gender-based violence and gender inequality'.[22]

2.27 eSafety's research indicates that women and their children who are experiencing domestic and family violence 'almost always experience technology-facilitated abuse designed to extend coercion and control over their lives'.[23] This was confirmed by the Women's Services Network

---

17 Office of the eSafety Commissioner, *Adult cyber abuse*, available at: https://www.esafety.gov.au/key-issues/adult-cyber-abuse (accessed 7 March 2022).

18 eSafety Commissioner, *Submission 53*, p. 14.

19 eSafety Commissioner, *Submission 53*, p. 14.

20 eSafety Commissioner, *Submission 53*, p. 21.

21 WESNET, *Submission 25*, p. 2.

22 WESNET, *Submission 25*, p. 2.

23 eSafety Commissioner, *Submission 53*, p. 29.

(WESNET), who reported that a 2020 study found that almost all survey participants had experienced technology-facilitated abuse.[24]

2.28 Methods of technology-facilitated abuse included text messaging, tracking apps, FaceTime and iCloud, and the misuse of government accounts such as MyGov.[25] WESNET also reported that the study indicated that women were regularly forced to film and record intimate images, suggesting that image-based abuse was being utilised by perpetrators.[26]

2.29 Australia's National Research Organisation for Women's Safety (ANROWS) outlined some of the other forms that technology-facilitated abuse can take:

> Some of the other interesting research that we've come across is when people are transferring money. The banks have reported that even the comments in bank transfers have been used to perpetuate abuse. So there are many ways of using technology to continue the abuse and that relationship. It could be using devices in toys to stalk or monitor but also using social media, pretending to be a friend, pretending to be someone else as a way of both targeting children and targeting, usually, the mother. So there's certainly a case for thinking about the threshold, because often it might not meet that threshold, and also about a more fulsome understanding by service providers of how technologies are used to perpetuate abuse.[27]

2.30 Evidence suggests that technology-facilitated abuse is markedly different from other forms of abusive behaviour due to the capacity for harm to be committed immediately, more easily and with greater reach.[28] Children are also either likely to be victims of technology-facilitated abuse, or used as a perpetrator's method of monitoring the victim and continuing the abuse.[29]

2.31 The impact of technology-facilitated abuse is all-encompassing for many victims, as ANROWS highlighted:

> The constant monitoring and abuse enacted through technology creates a sense of omnipresence for victims, making it feel as though they're constantly

---

24 WESNET, *Submission 25*, p. 3.

25 WESNET, *Submission 25*, p. 4.

26 WESNET, *Submission 25*, p. 4.

27 Ms Padma Raman, Chief Executive Officer, Australia's National Research Organisation for Women's Safety (ANROWS), *Committee Hansard*, 28 January 2022, p. 5.

28 Dr Bridget Harris, members of The Independent Collective of Survivors, Molly Dragiewicz and Delanie Woodlock, *Submission 17*, p. 2.

29 Dr Bridget Harris, members of The Independent Collective of Survivors, Molly Dragiewicz and Delanie Woodlock, *Submission 17*, p. 8.

> being watched by their perpetrators. Workers said this made victims hypervigilant and fearful and made them feel as if the abuse would never end or that they would never be able to escape. Technology-facilitated abuse has the ability to impact all facets of victims' lives. They don't feel safe at home, work, while studying or in social interactions.[30]

2.32 Witnesses also noted that technology-facilitated abuse is often part of a broader pattern of domestic and family violence and accordingly needed to be considered in the context of domestic and family violence more broadly. Ms Karen Bentley, CEO of WESNET, stated that:

> From our perspective, the problem that we've got is that for a survivor who's experiencing domestic and family violence or some other form of gender based violence the abuse that they experience online through technology is one aspect of a wide range of tactics that the abuser will use to control and coerce the victim. When we have approaches which try to just deal with one aspect, everything else moves on the other side, so it's vitally important that we have a much greater understanding about the impact and the dynamics of how domestic and family violence actually works. Our responses to it can't be the cybersafety training ones that we have for the average person who is trying to stop a nameless, faceless Russian hacker—for want of a better example—and to protect their online presence. These are very targeted. They have intimate knowledge of the victim and will use and target the victim very, very directly.[31]

2.33 WESNET further highlighted the immense impact of domestic violence perpetrators using online platforms to target or harass victims:

> …violence against women is not acceptable in any form on any platform and it has really devasting consequences online as well as in real life… people can be anonymous online and troll the hell out of people, and many of our survivors are badly affected by the fact that those people are doing this abuse anonymously and they can't be found and tracked.[32]

Image-based abuse

2.34 Image-based abuse, while considered a form of technology-facilitated abuse, is predominantly in relation to the sharing or threatened sharing of intimate images without consent. Otherwise known as 'revenge porn', intimate

---

30 Ms Padma Raman, ANROWS, *Committee Hansard*, 28 January 2022, p. 2.

31 Ms Karen Bentley, WESNET, *Committee Hansard*, 28 January 2022, p. 5.

32 Ms Karen Bentley, WESNET, *Committee Hansard*, 28 January 2022, pp 7-8.

images that fall under the definition provided in legislation include the exposure of:

> a person's genital area or anal area (whether bare or covered by underwear); a person's breasts (if the person identifies as female, transgender or intersex); private activity (for example, a person undressing, using the bathroom, showering, bathing or engaged in sexual activity); or a person without attire of religious or cultural significance if they would normally wear such attire in public.[33]

2.35 New technologies have further complicated image-based abuse due to the emergence of 'deepfake' technology. A 'deepfake' is a:

> digital photo, video or sound file of a real person that has been edited to create an extremely realistic but false depiction of them doing or saying something that they did not actually do or say.[34]

2.36 eSafety states that deepfakes have the potential for positive uses, such as for entertainment, education and medical purposes, but that such technology poses significant risks when used harmfully. eSafety's position statement on deepfakes explained that they can be used for a range of purposes that are damaging to victims, such as:

- Creating 'fake news' or hoaxes;
- Producing falsified pornography material;
- Stealing someone's identity or impersonating someone; and
- Extorting victims by creating fake material of them and then threatening to release it to their contacts.[35]

Online child sexual exploitation

2.37 Online child sexual exploitation includes a range of behaviour and offences which relate to the grooming and sexual abuse of children. Types of behaviour that fall under the definition of online child sexual exploitation include:

---

33 eSafety Commissioner, *Submission 53*, p. 20.

34 Office of the eSafety Commissioner, *Deepfake trends and challenges – position statement*, 23 January 2022, available at: https://www.esafety.gov.au/industry/tech-trends-and-challenges/deepfakes (accessed 6 March 2022).

35 Office of the eSafety Commissioner, *Deepfake trends and challenges – position statement*, 23 January 2022, available at: https://www.esafety.gov.au/industry/tech-trends-and-challenges/deepfakes (accessed 6 March 2022).

- An adult engaging a child in a chat about sexual acts
- An adult sending nude or pornographic images of themselves to a child or exposing themselves via live streaming
- Asking a child to view pornographic images/videos
- Asking a child to perform sexual acts, expose themselves or share a sexual image
- Taking or making and sharing or showing indecent images of children.[36]

Violent or abhorrent content

2.38 The production and publication of violent or abhorrent material online, particularly in relation to livestreaming, has been of legislative scrutiny since the Christchurch terrorist attack in March 2019. The legislative definition of such content states that abhorrent violent material is considered as:

> audio and/or visual content produced by a perpetrator or accomplice of a terrorist act involving serious physical harm or death, murder or attempted murder, torture, rape or kidnapping involving violence.[37]

2.39 The legislative definition contains a number of exceptions, such as bystander coverage, journalism, research and artistic purposes.[38]

2.40 eSafety states that such material poses significant risk to online users, including increased trauma and suffering to victims and their families, potential radicalisation and extremism, and the potential for these kinds of content to be used to incite fear.[39]

Promotion of harmful behaviours

2.41 Content produced online has the capacity to encourage or promote destructive or unhealthy behaviours for vulnerable users. eSafety points to the examples of self-harm, suicide and eating disorders as topics which fall under this category.[40]

---

36 Australian Centre to Counter Child Exploitation, *What is online child sexual exploitation?*, available at: https://www.accce.gov.au/help-and-support/what-is-online-child-exploitation (accessed 28 February 2022).

37 eSafety Commissioner, *Submission 53*, p. 23.

38 eSafety Commissioner, *Submission 53*, p. 23.

39 eSafety Commissioner, *Submission 53*, p. 24.

40 eSafety Commissioner, *Submission 53*, p. 26.

2.42 Suicidal ideation is a common topic of concern in relation to online harm. Orygen stated that social media 'has the potential for increasing the risk of contagion or copycat behaviours and sharing information about suicide methods'.[41]

2.43 Further, online content depicting eating disorders (such as anorexia and bulimia) in a positive light, promoting disordered behaviour, or providing instruction in disordered eating, can encourage eating disorders for vulnerable people.[42]

Community harm

2.44 Harm that impacts the community refers to situations where content online has the potential to cause harm, or causes actual harm, to the community. This form of harm encompasses situations such as:

- Inciting violence or hatred against particular groups, such as gender, racial or disability groups (also known as hate speech);
- Promoting or distributing material relating to extremism and terrorism, including live-streaming violent attacks; and
- Spreading misinformation, disinformation, or encouraging mistrust in government institutions.

2.45 Further, types of individual and economic harm can have broader social impacts which result in community harm. For example, CSAM impacts the individual victim, but also can impact the victims' family and support networks, and more broadly can cause fear, anxiety and anger in the local community.

Online hate

2.46 Online hate is a term that covers a range of extremely harmful practices that have been known to proliferate on social media platforms in addition to other digital arenas. While no legal standard exists for the definition of 'hate speech' in Australia,[43] eSafety defines online hate as 'any hateful posts about

41 Orygen, *Submission 27*, p. 3.

42 Eating Disorders Families Australia, *Submission 37*, p. 1.

43 Digital Industry Group Inc., *Submission 46*, p. 5.

a person or group based on their race, religion, ethnicity, sexual orientation, disability or gender'.[44] The range of forms that online hate can take include:

- Discrimination;
- Hate speech;
- Racism;
- Misogyny, misandry and other forms of sex-based discrimination;
- Sexual harassment; and
- Homophobia or other forms of discrimination based on sexual orientation.[45]

2.47 The eSafety Commissioner makes a distinction between discrimination as opposed to hate speech, noting that while discrimination can be targeted against an individual, hate speech targets an entire group.

2.48 Harmony Alliance stated that groups who 'do not have the same access to power and privilege as the dominant groups' are likely to experience discrimination in various forms in online settings.[46] They pointed to research conducted by eSafety which approximated that one in seven (approximately 14 per cent) adults online had been targeted by online hate speech on a social media platform in the period between August 2018 and August 2019.[47]

2.49 Ms Nyadol Nyuon, incoming Chair of Harmony Alliance, pointed to examples such as the targeting of African groups on social media platforms in 2016 which attracted nationalistic and neo-Nazi groups.[48] She also highlighted her own experiences of being racially abused, explaining how she had received comments 'calling for the culling of people who looked like me' and being attacked by a police officer who 'called me an ignorant C-word who should F-off back to the war-torn shithole country I came from'.[49]

2.50 Focusing particularly on the experiences of migrant and refugee women in Australia, Harmony Alliance stated that online discrimination can take many forms, such as:

---

44 eSafety, *Online Hate*, available at: https://www.esafety.gov.au/young-people/online-hate (accessed 7 March 2022).

45 eSafety Commissioner, *Submission 53*, p. 26.

46 Harmony Alliance, *Submission 34*, p. 2.

47 Harmony Alliance, *Submission 34*, p. 3.

48 Ms Nyadol Nyuon, Incoming Chair, Harmony Alliance, *Committee Hansard*, 22 December 2021, p. 8.

49 Ms Nyadol Nyuon, Harmony Alliance, *Committee Hansard*, 22 December 2021, p. 7.

- insulting, humiliating, demeaning or offensive comments – both directed towards them as individuals and to their communities;
- derogatory language;
- threats of sexual and physical violence;
- threats against children;
- death threats;
- online stalking;
- distributing personal contact details online (doxing); and
- image-based abuse (for example, the non-consensual sharing of intimate or false photos online).[50]

2.51 Harmony Alliance also noted that a complicating factor for migrant or refugee groups was that some forms of harassment or abuse were conducted in languages other than English, making it difficult to monitor and detect on platforms which are monitored primarily by English speakers. Reporting abuse was also said to be difficult due to moderators being unable to understand the language or the contextual background. [51]

2.52 Some witnesses pointed out that discrimination was often not isolated to a particular characteristic of a person, and could be in relation to a number of other identity-related factors, such as gender, religion and sexual orientation.[52]

2.53 Harmony Alliance also noted that certain groups had experienced heightened levels of discrimination and hate speech during the COVID-19 pandemic. They pointed to evidence which suggested that Asian-Australians, in addition to young migrant and refugee women, had experienced an increase in abuse and discrimination in online settings.[53]

Disinformation and misinformation

2.54 Disinformation and misinformation on social media platforms is an issue receiving growing attention.

---

50 Harmony Alliance, *Submission 34*, pp 2-3.

51 Harmony Alliance, *Submission 34*, p. 4.

52 Harmony Alliance, *Submission 34*, p. 3.

53 Harmony Alliance, *Submission 34*, p. 3.

2.55 What constitutes 'disinformation' and 'misinformation' is not universally agreed. The Digital Industry Group Inc. (DIGI) noted that there is currently no consensus across stakeholders as to what these terms should include.[54] The Centre for Digital Wellbeing (CDW) provided the following definitions:

> Misinformation is a term used to describe content or information that is false but was created or shared without the intent to cause harm. Misinformation is false or out-of-context information that is presented as fact, and can include made-up news articles, false information shared on social media platforms, doctored images and videos, and scam advertisements.
>
> Conversely, disinformation is the purposeful or deliberate creation and dissemination of false information with the intention to mislead or cause harm. Disinformation can take many forms. It can include false or fake news content or fake news sites, images or text that are altered or distorted, or videos or commentary that include elements of fact mixed with elements of falsehood or exaggeration. Disinformation can also include real material used within a context that presents a distorted view of reality, such as a clip of a speech that is given a new and false attribution of meaning. The amalgamation of false information with truth is a common tool used in disinformation campaigns and is highly effective as a tactic of influence.[55]

2.56 The CDW explained that disinformation and misinformation are spread using a number of technological tools, such as:

- Tools to automatically generate news articles, post on social media and engage with others (known as 'bots') which behave identically to humans online, making them difficult to detect;
- 'Bot farms', which are large groups of bots which are designed to work in a coordinated way to provide the appearance of truthfulness, persuasiveness and popularity, while also boosting the content's reach on algorithm-generated social media sites; and
- Troll farms, which are groups working together to upload content which is often 'inflammatory, divisive and false' to social media services.[56]

2.57 The Australian Communications and Media Authority stated that disinformation and misinformation have the capacity to significantly damage people on an individual level and collectively:

---

54 DIGI, *Submission 46*, p. 34.

55 Centre for Digital Wellbeing (CDW), *Submission 47*, p. 16.

56 CDW, *Submission 47*, p. 17.

> [m]isinformation can pose a risk to people's health and safety. We have seen this with misinformation about COVID-19 and 5G technology … [o]nline, there is such a large amount of information from different sources that it can be hard to know who or what to believe. It may not be clear where the information has come from, who wrote it, or when it was produced. When we share something online, we do not always stop to think whether it is true. Misinformation can be new, surprising, or emotive. This can make us more likely to share it and it can often spread faster than the facts.[57]

Extremism and terrorism

2.58 Submitters to the inquiry argued that the proliferation of social media services' algorithms which amplify extreme and sensationalist content has resulted in an increasing trend in society towards extremism.[58] While extremism and terrorism have existed outside the internet for centuries, social media and digital platforms have provided a vehicle for those with extreme beliefs or ideologies to meet and advertise their cause.

2.59 The Department of Home Affairs (Home Affairs) outlined its concerns in relation to the spread of violent extremism via social media platforms and digital services:

> Digital platforms present deep-seated challenges for Australia's national efforts to contest and prevent violent extremism. Increasingly, violent extremists from across the ideological spectrum seek to use online methods to spread extreme and harmful propaganda, seed division and recruit individuals.[59]

2.60 Home Affairs stated that social media services are regularly used as 'a major conduit for terrorism and violent extremist content', but that due to Australian and international law enforcement's efforts in removing content, such content is now shifting towards platforms which are less willing or able to remove it.[60]

2.61 The CDW provided an example of a recent study which highlighted concerns regarding social media services and their impact on extremism:

> In one internal Facebook study, a researcher created a Facebook account for a fictional 41-year-old conservative mother with an interest in 'young children,

---

57 Cited in Free TV Australia, *Submission 42*, p. 7.

58 CDW, *Submission 47*, p. 15.

59 Department of Home Affairs (Home Affairs), *Submission 40*, p. 10.

60 Home Affairs, *Submission 40*, p. 11.

> parenting, Christianity, Civics and Community'. After this fictional account liked memes and joined conservative groups on the first day, Facebook began recommending almost exclusively right-wing content on the second day. By the fifth day, it was recommending QAnon content and right-wing conspiracy theories. Facebook's internal research found similar effects for a fictional liberal user.[61]

2.62 Similar findings were also reflected in the New Zealand Royal Commission's Inquiry into the 2019 terrorist attack in Christchurch, which identified that extremists and terrorists were utilising social media platforms to find one another, share information and spread their ideologies.[62]

Economic harm

2.63 Economic harm describes a situation where a range of types of financial harm are experienced due to conduct online. This includes a range of situations, such as:

- Scams or frauds committed online;
- Ransomware and other forms of technology-facilitated harm;
- False advertisements or representations; and
- False reviews of businesses which can cause financial hardship.

2.64 This form of harm, while important, is not examined in this report.

Prevalence of online harm

2.65 Currently available statistics indicate that digital spaces are saturated with online harm, with everyday Australians experiencing the repercussions. The eSafety Commissioner provided the following statistics in relation to online harm:

- Research conducted in 2017 indicated that 11 per cent of Australians over 18 years of age have been the target of image-based abuse, the majority of which was directed towards women between the ages of 18 and 24 years of age.[63]

---

61 CDW, *Submission 47*, p. 15.

62 CDW, *Submission 47*, p. 15.

63 eSafety Commissioner, *Submission 53*, p. 21.

- Between August 2018 and August 2019, 67 per cent of Australian adults had a negative experience online, ranging from unwanted online contact, security breaches, and hate speech or abuse;[64]
- One in five young Australians have experienced cyberbullying behaviour and one in five Australian children or young people admit to cyberbullying behaviour;[65] and
- During the 2020-2021 reporting period, eSafety received in excess of 23,500 reports from the public in relation to illegal or restricted content online, the vast majority of which concerned CSAM.[66]

2.66 It is important to note that some of there statistics are at least five years old. Given the exponential rise in the number of social media and digital platforms and the increasing power of the existing companies, it is safe to assume that these statistics may be significantly out of date.

2.67 The level of awareness of social media and digital platforms regarding the full extent of online harm being caused is unclear. Many large social media platforms do not provide clear statistics of the level of harm present on their services. Twitter, for example, publishes regular Rules Enforcement Transparency Reports which detail the number and kind of breaches of its terms of service within a reporting period.[67] Nonetheless, the reports do not provide an overall understanding of the number of users who experience online abuse or give an indication of the proportion of this behaviour across its platform. They also only represent the rate of detected abuse, which suggests that the true rates of online abuse are significantly greater if they are not detected.

2.68 In terms of bullying and harassment on its services, Meta stated that its latest Community Standards Enforcement Report found that the rate of bullying and harassment was 0.14-1.15 per cent on Facebook and 0.05-0.06 on Instagram, which meant that this form of harm was 'seen between 14 and 15 times per every 10,000 views of content on Facebook, and between 5 and

---

64 eSafety Commissioner, *Adults' negative online experiences – eSafety research*, August 2020, available at: <https://www.esafety.gov.au/sites/default/files/2020-07/Adults%27%20negative%20online%20experiences.pdf> (accessed 7 February 2022), p. 4.

65 eSafety Commissioner, *Submission 53*, p. 15.

66 eSafety Commissioner, *Submission 53*, p. 18.

67 Twitter, *Rules Enforcement – Transparency Report, January to June 2021*, 25 January 2022, available at: https://transparency.twitter.com/en/reports/rules-enforcement.html#2021-jan-jun (accessed 23 February 2022).

6 times per 10,000 views of content on Instagram'.[68] It is unclear how Meta identified these statistics.

2.69 These statistics provided by Meta offer information on only one form of harm that is present on social media platforms. Further, these statistics minimise the level of harm being caused by focusing only on the rate of identified bullying and harassment by views, rather than the impact on the victim which can be disproportionately extreme.

Who is most at risk online?

2.70 Online harm can be experienced by any user of the internet. Individuals from different backgrounds provided evidence to the Committee in relation to their experiences of online abuse. From high-profile media personalities to children, online harm does not discriminate.

2.71 Notwithstanding this, it is widely recognised that certain groups are significantly more likely to experience online harm or are more vulnerable to the effects of dangerous behaviour online. This section describes these groups and how online harm affects them.

Children and young people

2.72 Children are widely recognised as amongst the most at-risk groups in relation to online harm. The harmful content that children and young people are exposed to is diverse, including (but not limited to):

- Online child sexual exploitation and children being contacted and groomed by abusive perpetrators;[69]
- Accessing inappropriate content beyond a child or young person's developmental level, which can lead to distress, desensitisation and other forms of harm (e.g. pornography, violence against people and animals, and self-generated sexual content);[70]
- Terrorist and other extremist content;[71]

---

68 Meta, *Submission 49*, p. 16. Meta noted in its submission that this statistic was reflective only of bullying and harassment where Meta did not need additional information to determine if it violated its policies, such as a report from a person experiencing the conduct.

69 eSafety Commissioner, *Submission 53*, p. 17.

70 eSafety Commissioner, *Submission 53*, p. 17 and 23.

71 Home Affairs, *Submission 40*.

- Disordered eating and body dysmorphia;[72]
- 'Sextortion', where a young person is coerced or willingly provides explicit images of themselves to another person, who then threatens to share the images with the victim's friends or family unless the victim provides more explicit images;[73]
- Discrimination, including racism, hate speech and homophobia;[74] and
- Cyber-bullying, harassment, stalking and other forms of harm aimed specifically at an individual user.[75]

2.73 Children and young people experience online harm at alarming rates. In recent research conducted by eSafety in August-September 2021, almost half of the surveyed children in its study had experienced hurtful or nasty treatment online within the past year, and one in ten children had been targeted with online hate speech. [76] The report found that children also routinely engage in risky behaviour online, finding that six in ten children had communicated with someone they first met online, one in eight children have sent a photo or video of themselves to a person they initially met online, and one in eight had met someone they had first met online.[77] These proportions were also noted by eSafety to be a significant increase compared to rates observed in 2016, where children's participation in risky activities was considerably lower.[78]

---

72 Butterfly Foundation, *Submission 10*; Eating Disorders Families Australia, *Submission 37.*

73 Ms Sonya Ryan, Chief Executive Officer and Founder, The Carly Ryan Foundation (CRF), *Committee Hansard*, 21 December 2021, p. 7.

74 Professor Amanda Third, Professorial Research Fellow, Young and Resilient Research Centre, *Committee Hansard*, 21 December 2021, p. 30; Ms Christine Morgan, NMHC, *Committee Hansard*, 21 January 2022, p. 5.

75 Ms Sonya Ryan, CRF, *Committee Hansard*, 21 December 2021, p. 1; eSafety Commissioner, *Submission 53*, p. 14.

76 Office of the eSafety Commissioner, *Mind the Gap – Parental awareness of children's exposure to risks online*, February 2022, available at: https://www.esafety.gov.au/sites/default/files/2022-02/Mind%20the%20Gap%20%20-%20Parental%20awareness%20of%20children%27s%20exposure%20to%20risks%20online%20-%20FINAL.pdf (accessed 8 February 2022), p. 7.

77 Office of the eSafety Commissioner, *Mind the Gap – Parental awareness of children's exposure to risks online*, February 2022, available at: https://www.esafety.gov.au/sites/default/files/2022-02/Mind%20the%20Gap%20%20-%20Parental%20awareness%20of%20children%27s%20exposure%20to%20risks%20online%20-%20FINAL.pdf (accessed 8 February 2022), p. 7.

78 Office of the eSafety Commissioner, *Mind the Gap – Parental awareness of children's exposure to risks online*, February 2022, available at: https://www.esafety.gov.au/sites/default/files/2022-

2.74 eSafety also examined the online behaviour of teenagers, finding that it was common for teenagers to be exposed to negative online content. The report stated:

- Almost two-thirds of young people aged 14–17 were exposed in the past year to negative content, such as content relating to drug taking, suicide or self-harm, or gory or violent material.
- Seven in ten young people aged 14–17 have seen sexual images online in the past year, while close to half have received sexual messages from someone online in the past year.[79]

2.75 eSafety's findings were corroborated by other research. One study found that more than 70 per cent of vulnerable children and youth have witnessed harmful content online, such as violent or explicit content.[80]

2.76 Children's safety organisations also reported that their activities indicated a high prevalence of exposure to online harm amongst children and young people. yourtown, the operators of Kids Helpline, explained that in 2020 approximately 4.5 per cent of calls it received included reference to cybersafety issues.[81] yourtown stated that online safety concerns were most prevalent for clients under the age of 18 years old, which were often accompanied by concerns relating to bullying, mental health, and suicide ideation.[82] From 2016 onwards, Kids Helpline has received approximately 209 contacts per year, particularly in the 13- to 18-year-old age group, in

02/Mind%20the%20Gap%20%20-%20Parental%20awareness%20of%20children%27s%20exposure%20to%20risks%20online%20-%20FINAL.pdf (accessed 8 February 2022), p. 40.

79 Office of the eSafety Commissioner, *Mind the Gap – Parental awareness of children's exposure to risks online*, February 2022, available at: https://www.esafety.gov.au/sites/default/files/2022-02/Mind%20the%20Gap%20%20-%20Parental%20awareness%20of%20children%27s%20exposure%20to%20risks%20online%20-%20FINAL.pdf (accessed 8 February 2022), p. 7.

80 The Social Switch Project and Dr Faith Gordon, *Online Harms Experienced by Child and Young People: 'Acceptable Use' and Regulation – Executive Summary*, November 2021, available at: https://static1.squarespace.com/static/5d7a0e7cb86e30669b46b052/t/618b7c55a660d050880bb03d/1636531286894/Online+Harms+Research+November+2021+-+Executive+Summary.pdf (accessed 6 February 2022), p. 2.

81 Ms Kathryn Mandla, Head, Advocacy and Research, yourtown, *Committee Hansard*, 21 December 2021, p. 32.

82 Ms Kathryn Mandla, yourtown, *Committee Hansard*, 21 December 2021, p. 32.

relation to online or texting-based sexual activity, including sexting and self-distribution of explicit images.[83]

2.77 The Carly Ryan Foundation (CRF) stated that twenty per cent of teenagers receive unwanted or inappropriate content, such as violent or sexual content, via online means.[84] Other groups working directly with young people, such as The Daniel Morcombe Foundation (DMF), stated that in their experience that there had been a 'definite increase' in the rate of technology-assisted harmful sexual behaviours online.[85]

2.78 Further, children are entering online spaces such as social media at younger ages than has previously been observed.[86] This may indicate that online harm could be starting earlier than previously observed.

2.79 The reasons why online harm presents as a uniquely dangerous threat to children and young people are complex. Ms Sonya Ryan, Chief Executive Officer (CEO) and Founder of the CRF, suggested that children and young people are innately willing to trust others and share information. Ms Ryan explained:

> They have insecurities, they're looking for validation and they want to be connected and be part of something. Often those vulnerabilities and the conditioning provided to them from what they're seeing through media and the online world sets them up, potentially, for an amount of suffering, whether that be physical, emotional or mental, because they simply cannot live up to what they're seeing around them in the online space. They often don't feel like they're enough. When they're looking for that validation, it leaves them very vulnerable to inappropriate content and contact.[87]

Experiences of managing young people's online behaviour

2.80 The Committee received evidence demonstrating that families are struggling to manage their children's online behaviours. eSafety's research found that parents were often unaware of the extent to which children and young people were accessing harmful content or experiencing online abuse, and

---

83 Ms Kathryn Mandla, yourtown, *Committee Hansard*, 21 December 2021, p. 32.

84 Ms Sonya Ryan, CRF, *Committee Hansard*, 21 December 2021, p. 2.

85 Ms Tracey McAsey, Manager, The Daniel Morcombe Foundation, *Committee Hansard*, 21 December 2021, p. 10.

86 Professor Amanda Third, Young and Resilient Research Centre, *Committee Hansard*, 21 December 2021, p. 26.

87 Ms Sonya Ryan, CRF, *Committee Hansard*, 21 December 2021, p. 3.

were often unaware of how significantly these experiences were impacting their children. Disturbingly, eSafety found that parents had a significantly lower awareness of their children's exposure to sexual material than the actual rate of children's exposure.[88]

2.81 Parents were reported to feel pressured into granting their children access to social media platforms due to their peers having access.[89] As the CDW noted:

> Parents and carers are in an unenviable situation when it comes to regulating their children's social media use. If they do not allow their children to use social media, their children may be excluded or socially isolated. However, allowing social media use may negatively affect their child's mental health.[90]

2.82 Further, anecdotal evidence from child safety organisations suggested that parents and carers 'frequently lack confidence in their ability to help children stay safe or to effectively deal with unsafe experiences', a finding which was supported by eSafety's research.[91] This was corroborated by witnesses, who stated that parents often felt overwhelmed by the pace of technological development and 'simply give up', which can result in children 'tak[ing] advantage of their parent's limited focus, lack of tech awareness and lack of time'.[92]

2.83 Witnesses also noted that while there is software and technological tools for parents to monitor and control their children's access to online services, the software could be 'expensive, with monthly fees and occasionally bugs', in addition to issues for parents such as time poverty, and technological illiteracy to ensure that the most appropriate software is used.[93] Further,

---

88 Office of the eSafety Commissioner, *Mind the Gap – Parental awareness of children's exposure to risks online*, February 2022, available at: https://www.esafety.gov.au/sites/default/files/2022-02/Mind%20the%20Gap%20%20-%20Parental%20awareness%20of%20children%27s%20exposure%20to%20risks%20online%20-%20FINAL.pdf (accessed 8 February 2022), p. 9.

89 Ms Sonya Ryan, CRF, *Committee Hansard*, 21 December 2021, p. 3.

90 CDW, *Submission 47*, pp 10-11.

91 Body Safety Australia, *Submission 59*, pp 3-5.

92 The Synod of Victoria and Tasmania, Uniting Church in Australia, *Submission 52*, p. 43.

93 The Synod of Victoria and Tasmania, Uniting Church in Australia, *Submission 52*, p. 43.

parents could potentially put too much trust in the software to adequately protect children and relax their vigilance.[94]

2.84 Schools were also identified as being recipients and managers of online harm issues, with varying degrees of success. eSafety stated that it had found that cyberbullying and other forms of online harm between young people often have roots in the schoolground, which sometimes results in a platform moderator being unable to understand the context of the online abuse.[95]

2.85 The AMF stated that schools receive complaints from parents in relation to potential cyberbullying incidents between students, and are often expected to intervene in these situations.[96] The AMF suggested that this places significant pressure on school staff, who are expected to resolve the situation effectively.[97]

2.86 Body Safety Australia noted that schools experience issues when managing serious complaints between students, explaining that reporting potential offences can be complicated by the schools' duty of care not just to the victim but also the offender.[98] Body Safety Australia stated:

> Victims who report to schools and see little or no response are re-traumatised by the lack of action as well as the necessity of facing their abuser every day in class. We have seen victims of online and offline abuse being unable to continue their education or being forced to change schools because of the effects and the inaction by schools. In many cases, online abuse will follow them to new schools, resulting in ongoing harm to their education and well-being.[99]

2.87 Box 2.1 provides a case study of harm in relation to children in online settings, particularly in how social media platforms respond to concerns regarding safety.

Box 2.1 Case study: YouTube videos of young girls

94 The Synod of Victoria and Tasmania, Uniting Church in Australia, *Submission 52*, p. 43.

95 Mr Toby Dagg, Executive Manager, Investigations, Office of the eSafety Commissioner, *Committee Hansard*, 3 February 2022, p. 15.

96 Ms Sarah Davies, AMF, *Committee Hansard*, 21 December 2021, p. 19.

97 Ms Sarah Davies, AMF, *Committee Hansard*, 21 December 2021, p. 19.

98 Body Safety Australia, *Submission 59*, p. 5.

99 Body Safety Australia, *Submission 59*, p. 5.

Dr Michael Salter, an expert in child sexual exploitation and online abuse, provided an example of a YouTube video that had 'gone viral'. In the video, a blogger described how he had identified a concerning trend about the platform and its use of algorithms.[100]

The video stated that children (or their parents or carers) were uploading innocuous videos of the children doing things such as performing gymnastics or dancing. These videos had attracted communities of paedophiles who had found the videos via search terms relating to children, and had then placed timestamps on them to enable other abusers to find the most explicit parts of a video (e.g. where a child inadvertently exposes a body part). The comments on these videos were highly explicit and enabled paedophiles to connect with one another. Further, due to YouTube's algorithms which recommend material based on a user's search and watch history, the platform recommended similar videos to users, essentially creating an 'alternate' side to the platform which enabled paedophiles to easily access more material.[101]

After it was alerted to this situation, the platform's response was to de-monetise all videos on YouTube and turn off all comments in videos depicting children, which Dr Salter described as 'just a mass and very blunt intervention into the problem'.[102] YouTube (administered by Google) deleted the comments and accounts on the specific video, and removed thousands of videos containing inappropriate images of young people and channels containing inappropriate content.[103] Nonetheless, the platform declined to deactivate the algorithm-driven 'recommendations' system for this type of content due to the potential

100 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 11. The video cited by Dr Salter is available at <https://www.youtube.com/watch?v=O13G5A5w5P0>.

101 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 11.

102 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 11.

103 Daisuke Wakabayashi and Sapna Maheshwari, 'Advertisers Boycott YouTube After Pedophiles Swarm Comments on Videos of Children', *The New York Times*, 20 February 2019, available at: https://www.nytimes.com/2019/02/20/technology/youtube-pedophiles.html?action=click&module=RelatedCoverage&pgtype=Article&region=Footer (accessed 6 February 2022).

impact on content-producers who rely on the recommendations system for viewership.[104]

Women

2.88 Women are highly likely to be targeted for online abuse. eSafety's research indicates that women and girls 'face disproportionate levels of online abuse that is sexualised and violent', making up two-thirds of complaints made to eSafety regarding cyberbullying, image-based abuse and adult cyber abuse.[105]

2.89 eSafety research about women in the workplace identified that women experience a range of online abuse, including:

- Unwanted private messages;
- Negative comments about their content;
- Bullying or trolling;
- Defamatory comments;
- Offensive comments about race, ethnicity or gender;
- Receiving slurs against their professional name;
- Lies or rumours;
- Stalking;
- Impersonation or fake accounts;
- Threats of real-life harm or abuse; and
- Being the target of an 'anti' or 'hate' group.[106]

2.90 Women are more likely to experience online harm, particularly that which is gender-based in nature, such as image-based abuse, sexist and misogynistic harassment and abuse (including harassment involving appearance, virtue and fertility), and technology-facilitated abuse.[107] Women are also likely to

---

104 Max Fisher and Amanda Taub, 'On YouTube's Digital Playground, an Open Gate for Pedophiles', *The New York Times*, 3 June 2019, available at: https://www.nytimes.com/2019/06/03/world/americas/youtube-pedophiles.html (accessed 6 February 2022).

105 eSafety Commissioner, *Submission 53*, p. 29.

106 Office of the eSafety Commissioner, *Women in The Spotlight: How online abuse impacts women in their working lives*, 2021, available at: https://www.esafety.gov.au/research/how-online-abuse-impacts-women-working-lives (accessed 27 January 2022).

107 Ms Julie Inman Grant, Commissioner, Office of the eSafety Commissioner, *Committee Hansard*, 3 February 2022, pp 23-24.

receive violent threats online, including generalised violence, rape and murder.[108]

2.91 Certain groups of women were stated by witnesses to be more likely to be subject to online abuse. Ms Nicole Shackleton stated that, according to her research, there were four subgroups of women which appeared to be more likely to receive abuse:

1 they occupy positions of power (such as politicians);

2 they draw attention to the ways that women's experiences are rooted in systemic inequalities or they call out men, masculinity and the patriarchy for contributing to and benefiting from social injustices and inequalities (such as activists and some journalists);

3 they assert their right to occupy and participate in public spaces, particularly when they enter traditionally masculine spaces (such as women athletes and journalists); and

4 they step out of traditional gender roles, or they do not confirm to traditional ideas of femininity and beauty standards.[109]

2.92 Ms Sall Grover, founder and CEO of the app Giggle For Girls, outlined to the Committee her experiences of online abuse and how social media platforms have dealt with these instances:

> One of the most horrific images I've ever received—my last name is Grover, which is obviously like a muppet from Sesame Street. Since I was born I've had a little Grover toy, and one person would send me images every day of Grover in a noose. It was just one of those things. Because I was born early, I was in an incubator, and they put a Grover toy in there to be my protector, so I've always had that image of it. Getting this image of Grover in a noose affected me more than just the words. For example, I got one the other day: 'I hope you burn in hell.' You get to a point where you have to develop a thick skin so it's water off a duck's back, but you do, at the same time, internalise it. As with every other piece of abuse I've received and with this muppet in a noose, I reported it to Twitter every time it happened. They always came back saying it has not violated their terms and conditions. I was like: How? At what point are your terms and conditions violated, because it seems to me that someone can go and post any kind of abuse that they want but when your focus maybe is on, say, misinformation and you'll ban people for that—which is an issue of itself; I understand—but abuse is also an issue, and it's part of the enjoyment of using a service. So they've done nothing. I've never, ever, ever

108 Miss Sall Grover, Founder and Chief Executive Officer, Giggle for Girls Pty Ltd (Giggle), *Committee Hansard*, 28 January 2022, p. 3.

109 Ms Nicole Shackleton, *Submission 28,* p. 9.

> had somebody removed or punished from Twitter for sending death threats or rape threats—ever.[110]

2.93 Women are also likely to be targeted in their workplace or in connection with their work. According to eSafety research, 35 per cent of women have experienced online abuse in relation to or as a consequence of their work.[111] Of this cohort, certain groups were more likely to experience online abuse, such as those with an online or media public profile, those living with a disability, those identifying as LGBTIQA+, or those aged between 18 to 34 years old.[112] Concerningly, women reported receiving abuse at work (such as via a work email) or by colleagues in the same industry.[113] This had a detrimental impact on victims' careers, such as feeling unsafe and less able at their jobs, reducing (temporarily or permanently) online activity, shying away from or declining leadership positions, or leaving their job or industry.[114]

Women in prominent positions

2.94 Women in public or prominent positions, such as journalists, politicians, sportswomen, and other public figures, have been identified as a particular group that experience higher levels of online abuse (see Box 2.2). The eSafety Commissioner stated that women in the public eye are recognised as receiving extreme forms of abuse:

> … we have a social media self-defence program called Women in the Spotlight, which is targeting women who are politicians, journalists and in the public eye, because it's not just greater prevalence; it's the way that the content

---

110 Miss Sall Grover, Giggle, *Committee Hansard*, 28 January 2022, p. 4.

111 Office of the eSafety Commissioner, *Women in The Spotlight: How online abuse impacts women in their working lives*, 2021, available at: https://www.esafety.gov.au/research/how-online-abuse-impacts-women-working-lives (accessed 27 January 2022).

112 Office of the eSafety Commissioner, *Women in The Spotlight: How online abuse impacts women in their working lives*, 2021, available at: https://www.esafety.gov.au/research/how-online-abuse-impacts-women-working-lives (accessed 27 January 2022).

113 Office of the eSafety Commissioner, *Women in The Spotlight: How online abuse impacts women in their working lives*, 2021, available at: https://www.esafety.gov.au/research/how-online-abuse-impacts-women-working-lives (accessed 27 January 2022).

114 Office of the eSafety Commissioner, *Women in The Spotlight: How online abuse impacts women in their working lives*, 2021, available at: https://www.esafety.gov.au/research/how-online-abuse-impacts-women-working-lives (accessed 27 January 2022).

> manifests. It's sexualised, it's violent; it's about appearance, supposed virtue, fertility. It's designed to humiliate and silence.[115]

2.95 Concurring with this point, Ms Nicole Shackleton stated that women in prominent positions are likely to experience gendered abuse on online forums, noting examples such former Prime Minister Julia Gillard, Victorian MP Fiona Pattern, and Federal Senator Dr Mehreen Faruqi as women in politics who have spoken out about the online abuse they have experienced.[116]

2.96 Ms Nicolle Flint MP, Member for Boothby, provided a submission to the Committee outlining her experiences of receiving gendered online abuse in addition to being stalked online and offline.[117] She also pointed to other examples of women in prominent positions who had experienced abuse online, including journalists Ms Leigh Sales and Ms Van Badham, former Federal Minister the Hon. Kate Ellis, and wife of the current Prime Minister, Ms Jenny Morrison.[118]

2.97 Dr Kate Hall, Head of Mental Health and Wellbeing at the Australian Football League (AFL), suggested that both men and women in the public view are routinely objectified and dehumanised 'because people project all of their own desires and wants on that individual', which perpetuates abuse.[119] Further, she stated that people in prominent positions often feel that they must accept the abuse and 'toughen up' as it is part of their job, which Dr Hall stated was psychologically unsafe and contributed to further trauma.[120]

2.98 Ms Erin Molan provided her experiences of online abuse to the Committee. She explained that, in her position as a sports journalist on *The Footy Show*, she had been subject to online abuse. She detailed abuse such as receiving negative and abusive comments via her social media feeds, and threatening direct messages, including one message where the person hoped to hurt her

---

115 Ms Julie Inman Grant, Commissioner, Office of the eSafety Commissioner, *Committee Hansard*, 3 February 2022, p. 24.

116 Ms Nicole Shackleton, *Submission 28*, p. 10.

117 Ms Nicolle Flint MP, *Submission 70*, pp 7-8.

118 Ms Nicolle Flint MP, *Submission 70*, pp 4-6.

119 Dr Kate Hall, Head of Mental Health and Wellbeing, Australian Football League (AFL), *Committee Hansard*, 1 February 2022, p. 18.

120 Dr Kate Hall, AFL, *Committee Hansard*, 1 February 2022, p. 18.

and her then-unborn child. These messages made Ms Molan fear for her safety and that of her family, but she did not speak publicly due to shame and embarrassment.[121]

2.99 Ms Molan approached the social media companies involved to attempt to stop the abuse but found it difficult to have her concerns taken seriously or for the abuse to be addressed. In one instance she reported a user who had threatened to try and 'kill the child within my stomach, and they came back and said that it didn't meet the threshold for inappropriate behaviour'.[122]

2.100 See Box 2.2 for a case study relating to online abuse targeting women in prominent positions.

Box 2.2 Case Study: Ms Tayla Harris

Ms Harris is a prominent sportswoman for the Australian Football League Women's (AFLW). Ms Harris has been subject to online abuse following the publication of a photo of her playing football in 2019, which she dubbed 'kicking-photogate'. [123] The photo depicted Ms Harris performing a follow-through of a kick, in a classic AFL pose.

Ms Harris stated that, following the publication of the picture, she received sexualised and disturbing comments, tags and direct messages from largely anonymous accounts. She received insults based on her personal character, which she found distressing. Ms Harris also stated that this abuse was what she described as 'a pile-on', otherwise known as a volumetric attack.[124]

Ms Harris' story went 'viral' around the world, and she continued to make comments in the public sphere about the nature of the abuse she was receiving, which further increased the attacks on her. Ms Harris expressed the view that some comments were aimed at silencing her.[125] In regard to having the harmful material taken down, Ms Harris stated that she attempted to report the content to the platforms but found the process very difficult. She also expressed that she had hoped that live

121 Ms Erin Molan, *Committee Hansard*, 18 January 2022, pp 2-3.

122 Ms Erin Molan, *Committee Hansard*, 18 January 2022, p. 2.

123 Ms Tayla Harris, Australian Football League Women's (AFLW), *Committee Hansard*, 1 February 2022, p. 20.

124 Ms Tayla Harris, AFLW, *Committee Hansard*, 1 February 2022, p. 20.

125 Ms Tayla Harris, AFLW, *Committee Hansard*, 1 February 2022, p. 20.

sports broadcasters or the AFLW would be able to moderate content and report abuse as it arose but recognised that it would require a non-stop effort to do so.[126]

Culturally and linguistically diverse people

2.101 People from culturally and linguistically diverse backgrounds (otherwise known as CALD communities) experience higher rates of online abuse than average, particularly in regards to hate speech and extremism. eSafety reports that eighteen per cent of people from CALD communities experience harm, compared to the national average of fourteen per cent.[127]

2.102 The Committee received powerful testimony from Ms Nyadol Nyuon, Incoming Director of the Sir Zelman Cowen Centre and Chair of Harmony Alliance, who detailed her personal experiences online. Ms Nyuon stated:

> The first online attack I received came after my first-ever media appearances on national television. The abuse was predominantly racist in nature, and some of the abuse used such violent language, including calling for the culling of people who look like me. I remember taking screenshots of the pictures of some of the individuals who directed the worst abuse, hoping that, at the very least, I might avoid them in public.
>
> The second attack was more sustained and reached every presence I had online. In what the eSafety Commissioner described at a Senate hearing as 'volumetric attack', I was tracked across all social media platforms and trolled predominantly with racist abuse. One came from a police officer, who called me an ignorant C-word who should F-off back to the war-torn shithole country I came from. He later apologised, and I accepted his apology. This time, though, the abuse and many things that were happening made me take three months off from work. The online abuse was not the only reason, but it played a substantial role in me taking the time to literally try to heal and reconnect again with a sense of safety. Because of that, I no longer share pictures of my children online, I prefer that my family members do not follow

126 Ms Tayla Harris, AFLW, *Committee Hansard*, 1 February 2022, p. 20.

127 eSafety Commissioner, *Submission 53*, p. 29.

> me online so they do not receive abuse, and I am constantly on watch to remove abuse that pops up on almost a daily basis.[128]

2.103 In addition to her personal experiences online, Ms Nyuon pointed to evidence suggesting that migrant and refugee women are at risk for online facilitated abuse, particularly in the context of family violence. She cited examples she had witnessed where men 'continue their campaign of terror on women by abusing them online, sharing their personal images without consent'.[129] Complicating factors in these situations include that these women may not speak English proficiently or at all, which means finding recourse for the abuse is limited. Further, the abuse can be conducted in languages other than English, which Ms Nyuon stated meant that hosting platforms often could not assist in the removal of the content because they did not understand the language being used.[130]

2.104 Further, Ms Nyuon suggested that social media provides an outlet for people to air 'racism, racist bullying and discrimination based on colour'.[131] She pointed to an example in 2016 where nationalistic and neo-Nazi groups used social media to issue threats towards African communities in Melbourne, and utilised the platforms to recruit people to do 'night work' to attack African youths.[132]

People living with disability or medical conditions

2.105 People living with disability or particular medical conditions are at higher risk of abuse online, which often tends to focus on their disability and/or their physical appearance.[133]

2.106 Ms Carly Findlay OAM outlined her experiences to the Committee, stating that she has had an active online presence since 1996 and has benefitted from this use, including through friendship, work opportunities and networking with the disability and facial difference communities. Ms Findlay expressed the view that for those in the disability community or

---

128 Ms Nyadol Nyuon, Incoming Director, Sir Zelman Cowen Centre, Victoria University; Chair, Harmony Alliance, *Committee Hansard*, 22 December 2021, p. 7.

129 Ms Nyadol Nyuon, Harmony Alliance, *Committee Hansard*, 22 December 2021, p. 7.

130 Ms Nyadol Nyuon, Harmony Alliance, *Committee Hansard*, 22 December 2021, p. 7.

131 Ms Nyadol Nyuon, Harmony Alliance, *Committee Hansard*, 22 December 2021, p. 8.

132 Ms Nyadol Nyuon, Harmony Alliance, *Committee Hansard*, 22 December 2021, p. 8.

133 eSafety Commissioner, *Submission 53*, p. 29.

other marginalised groups in society, the internet could be a place of safety and community.[134]

2.107 Ms Findlay had been wary of putting her photograph on the internet, as she has a rare facial difference condition called ichthyosis and had been subject to ridicule in offline settings.[135] She stated that after she published her photograph for work-related purposes, it was then repurposed by other online users to mock and abuse her:

> In December 2013 … I woke up to my photo being misused on Reddit. Reddit is a horrible cesspit of the internet … My photo was used on the 'what the fuck' forum. They were asking what the fuck had happened to my face. There were about 500 comments when I woke up, and they were all hideous. They were like, 'She looks like a glazed doughnut,' 'She looks like a lobster,' 'She looks like something my dog vomited up.' I sort of had a feeling that this would happen. I had a feeling my photo would be misused like this.[136]

2.108 Ms Findlay stated that she was able to satisfactorily manage the abuse herself after writing a Facebook post then publishing it as a direct response on Reddit, which 'changed the conversation'.[137] She reported other forms of abuse she had experienced, which included being stalked by a person she had engaged with online, a fake Instagram account being created with the intent of mocking her, and death threats.[138]

2.109 Ms Findlay also explained to the Committee that her interactions with digital platforms have given her the sense that the disability community is not supported. She explained that persons with facial differences often had content warnings applied to their photographs, which she attributed to platforms' artificial intelligence systems automatically applying it.[139] Ms Findlay also noted that it was very difficult for people with disabilities to be 'verified' by Twitter, as many of the requirements to have a verified account were 'quite prohibitive in an ableist world'.[140]

---

134 Ms Carly Findlay AO, *Committee Hansard*, 22 December 2021, p. 1.

135 Ms Carly Findlay AO, *Committee Hansard*, 22 December 2021, p. 1.

136 Ms Carly Findlay AO, *Committee Hansard*, 22 December 2021, p. 1.

137 Ms Carly Findlay AO, *Committee Hansard*, 22 December 2021, p. 1.

138 Ms Carly Findlay AO, *Committee Hansard*, 22 December 2021, p. 2 and 6.

139 Ms Carly Findlay AO, *Committee Hansard*, 22 December 2021, p. 5.

140 Ms Carly Findlay AO, *Committee Hansard*, 22 December 2021, p. 4.

2.110 While people living with disability are at risk online like most who utilise digital platforms, technology can be weaponised against them in particular ways. eSafety research suggests that women who live with disability are at higher risk for technology-facilitated abuse and in particular ways. These ranged from online harassment, to misusing their social media accounts, being monitor via spyware and other tracking technology, and image-based abuse. eSafety also found that perpetrators were often those who were closest to them, such as a partner or former partner, family members and carers.[141]

Aboriginal and Torres Strait Islander peoples

2.111 Aboriginal and Torres Strait Islander peoples experience hate speech on digital platforms at over three times the average rate for Australians online.[142] eSafety states that Indigenous women in particular are more likely than the general population to experience technology-facilitated abuse, but that Indigenous women in remote and regional communities are less likely to be aware of the issue.[143]

2.112 Mr Chad Wingard, football player in the AFL, explained that the abuse he received online had a substantial impact on him, and outlined his approach to dealing with it:

> If you're an Indigenous person or a person of colour or it's your sexuality or whatever it is you're being bullied about—I can only speak for being an Aboriginal person and a person of colour. However, my experience so far is that it takes a toll. It's draining and you think you'll let it slide or it's not the one that you think you need to call out. For me calling it out recently is because it affected me but not enough for me to give that person the limelight. It came to a point where I said, 'No, this is not on. I'm going to call out every single thing that happens now.' This is purely because I might be strong enough and have enough support around me to get through this, but I don't

141 eSafety Commissioner, *Technology-facilitated abuse of women with intellectual or cognitive disability*, August 2021, available at: https://www.esafety.gov.au/research/technology-facilitated-abuse-women-intellectual-or-cognitive-disability (accessed 7 February 2022).

142 Ms Julie Inman Grant, eSafety Commissioner, eSafety, *Committee Hansard,* 3 February 2022, p. 23.

143 eSafety, *Technology-facilitated abuse among Aboriginal and Torres Strait Islander women*, August 2021, available at: https://www.esafety.gov.au/research/technology-facilitated-abuse-among-aboriginal-and-torres-strait-islander-women (accessed 7 March 2022).

want 19-year-old kids coming from all over Australia who aren't capable and should not have to deal with this to even give these guys a chance.[144]

Other vulnerable groups

2.113 eSafety research suggests that other social groups are at higher risk of online harm. A non-exhaustive list of these groups include:

- People who identify as LGBTIQA+ or gender-divergent, with rates similar to those experienced by Aboriginal and Torres Strait Islander peoples;[145]
- People with particular religious beliefs;[146] and
- Older Australians.[147]

Repercussions of harms experienced online

2.114 Harm experienced by individuals online is not isolated to the internet. Online harm can have wide-reaching consequences which can impact a person's life in a number of ways. This section outlines the ramifications of online harm on individuals and the broader community.

Range of impacts caused by online harm

2.115 Online harm has the potential to cause significant and diverse forms of harm to individuals, both in online and offline environments. eSafety suggests that the impacts of online harm can include:

- Personal safety impacts – fear of psychological violence, physical violence and murder;
- Emotional and social impacts – annoyance, anger, humiliation, shame, guilt, self-blame, deception, betrayal and/or fear;
- Financial impacts – ability to work and earn an income, loss of financial security, restricted access to or knowledge of personal finances; and

144 Mr Chad Wingard, AFL, *Committee Hansard*, 1 February 2022, pp 24-25.

145 eSafety Commissioner, *Submission 53*, p. 29.

146 Ms Rita Jabri-Markwell, Adviser, Australian Muslim Advocacy Network, *Committee Hansard*, 1 February 2022, p. 21; Dr Andre Oboler, Chief Executive Officer and Managing Director, Online Hate Prevention Institute; Mr Peter Wertheim, Co-Chief Executive Officer, Executive Council of Australian Jewry, *Committee Hansard*, 22 December 2021, pp 12-13.

147 eSafety Commissioner, *Submission 53*, p. 29.

- Health and wellbeing impacts – anxiety, aggression, depression, self-destructive behaviour, physical health problems, intimate relationship difficulties, re-victimisation, disassociation, loss of self-esteem and confidence, withdrawal from social activities, lack of trust, substance abuse, ongoing trauma, self-harm and suicide.[148]

2.116 Other impacts resulting from negative online incidents experienced by adults include mental or emotional stress, or reputational damage.[149]

2.117 The impacts of online harm are also dependent on the type of harm experienced. For example, eSafety's research indicates that victims who have experienced image-based abuse felt a range of emotions and impacts, finding that:

> 65% felt annoyed, 64% felt angry, 55% felt humiliated, 40% felt depressed and 32% felt afraid for their safety. It negatively affected the self-esteem of 42%, the mental health of 41% and the physical wellbeing of 33% of victims.[150]

Trauma

2.118 Victims of online abuse can experience trauma from their experiences, particularly in relation to the most serious forms of online harm such as online child exploitation. Symptoms of trauma include 'fear, sleeplessness, paranoia, feelings of threat and lack of safety, and ostracism or social exclusion'.[151]

2.119 Dr Michael Salter explained that victims of online abuse can often experience trauma due to their experiences, depending on the type of harm and how long the harm continued for.[152] He stated that trauma refers to a psychological injury, which can be in relation to a one-off incident or continuous or repeated harm; some forms of trauma, such as post-traumatic stress disorder, can result in 'intrusive psychological symptoms' such as nightmares or flashbacks, which would ultimately resolve over time with psychological treatment.[153]

---

148 eSafety Commissioner, *Submission 53*, p. 27.

149 eSafety Commissioner, *Submission 53*, p. 28.

150 eSafety Commissioner, *Submission 53*, p. 27.

151 Ms Nicole Shackleton, *Submission 28*, p. 38.

152 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 14.

153 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 14.

2.120 Complex trauma, however, stems from 'repeated betrayal and violation', potentially involving violence, psychological and bodily invasion, and degradation.[154] Dr Salter stated that complex trauma is often experienced for longer time periods, particularly where a victim is young and the trauma has been committed regularly and over time.[155] For victims of child sexual exploitation, for example, Dr Salter stated that the trauma they experience is often lifelong, impacting on a victim's sense of safety, ability to have safe relationships, and their mental and psychosocial wellbeing.[156]

2.121 Dr Salter further explained the application of complex trauma to victims experiencing online harm, and for children and young people in particular the harm can be overwhelming and all-consuming:

> In terms of its link with online abuse, it's very typical that complex trauma is present for victims and survivors of online sexual exploitation for a range of reasons. There may have been offline abuse that then goes online, or there may have been online abuse—the child may have been induced into creating nude or sexual content. The continuing circulation of that material is extremely anxiety provoking and fear provoking. It's quite common for victims of online exploitation that they may be contacted repeatedly by abusers, who may in fact blackmail them and extort them with the content. It may be the same abuser or a different abuser. The fact of their online abuse may then become known to their peers, for example, at which point they may be subject to extensive bullying at school. There really can be the perception for this group that their life is destroyed. This is incredibly distressing for the young person, obviously, but also this anxiety, this trauma interferes with what we might say is a normal developmental pathway, their psychological but also physiological pathway. It interrupts psychological development and it interrupts neurological development; this then increases their risk of psychiatric and also autoimmune and other issues in adulthood.[157]

2.122 Mental health professionals told the Committee that they had witnessed the impact of trauma on victims of online abuse. Dr Kate Hall, AFL, stated that she had worked with a number of players who had demonstrated symptoms consistent with psychological trauma as a result of online abuse.[158]

---

154 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 14.

155 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 14.

156 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 9.

157 Dr Michael Salter, *Committee Hansard*, 18 January 2022, pp 14-15.

158 Dr Kate Hall, AFL, *Committee Hansard*, 1 February 2022, p. 10.

Mental health

2.123 Online harm can impact mental health for victims, with increased risks for depression and anxiety commonly cited as harmful repercussions of online abuse.[159]

2.124 On one hand, evidence suggested that children and young people are increasingly at risk of experiencing mental health issues. A 2016 study by Mission Australia and the Black Dog Institute examining youth mental health over a five year period found that, since 2012, one in four young people in Australia is at risk of serious mental illness, representing a 4.1 per cent increase over the reporting period.[160] Importantly, this study's estimations may be considerably out of date given the impact of the increased presence of online platforms in young people's lives since 2016.

2.125 The connection between the mental health of children and young people and their online activities, however, is yet to be made definitively. Notwithstanding the reported increase in demand for mental health services, the causal link between online harm and mental health is not clear. eSafety stated that it was critical to be cautious in drawing connections between online behaviour and users' mental health:

> It is important to take a nuanced and balanced view of children's and young people's experiences online and avoid drawing causal lines where they are not supported by evidence. The evidence before us suggests the relationship between mental health issues and social media use is complex. In fact, some usage can be positive and beneficial to mental health and wellbeing, while other usage patterns and experiences can be harmful.[161]

2.126 Professor Amanda Third noted that she had seen anecdotal evidence of extreme pressure on organisations that provide services to children, which was corroborated by groups such as the DMF and yourtown.[162] She posited, however, that there were numerous factors that could be attributed to this rise that were external to online matters, including unemployment, income

---

159 Dr Kate Hall, AFL, *Committee Hansard*, 1 February 2022, p. 11.

160 Mission Australia and The Black Dog Institute, *Youth mental health report: Youth Survey 2012-16*, 2016, available at: https://www.missionaustralia.com.au/publications/youth-survey/706-five-year-mental-health-youth-report/file (accessed 13 January 2022).

161 eSafety Commissioner, *Submission 53*, p. 30.

162 Professor Amanda Third, Young and Resilient Research Centre, *Committee Hansard*, 21 December 2021, p. 29.

difficulties, or unequal access to education. [163] eSafety agreed with this observation, noting that a broader contextual understanding of young people's lives was required in understanding mental health challenges:

> Most research exploring the intersection between social media and mental health notes there are mediating factors ranging from personality, underlying mental health issues, age, gender, socio-economic background, ethnicity, level of parental engagement, and a person's level of self-regulation in social media use.[164]

2.127 This point was also made by the NMHC, which stated that while mental health for young people was clearly in decline, the link to online usage was less clear. The NMHC stated that its work had identified that the 'trajectory is that mental health and wellbeing of our young people has been declining steadily for the last few years', beginning prior to, and accentuated by, the COVID-19 pandemic.[165] In conducting studies on youth mental health, the NMHC explained that it had consulted with youth advocates and its technical advisory group in relation to social media and online usage and the extent to which it impacts on mental health. Its findings indicated that 'it's probably an amplifier, not a driver in and of itself'.[166]

Harm to psychological and physical development in children and young people

2.128 Witnesses reported that there are significant repercussions to children and young people's exposure to unsafe or harmful content. Young people in particular are at risk of believing that the content they see online is representative of real life.

2.129 Ms Sonya Ryan, CEO of the CRF, stated that students today watch adult pornography to learn about sexual health and behaviour, which can lead to their believing that the sometimes violent or extreme content portrayed is normal. Ms Ryan stated that medical professionals have reported to the CRF that there are increasing rates of young people presenting with serious

---

163 Professor Amanda Third, Young and Resilient Research Centre, *Committee Hansard*, 21 December 2021, p. 29.

164 eSafety Commissioner, *Submission 53*, p. 30.

165 Ms Christine Morgan, NMHC, *Committee Hansard*, 21 January 2022, p. 9.

166 Ms Christine Morgan, NMHC, *Committee Hansard*, 21 January 2022, p. 9.

injuries, resulting from young people feeling 'pressured into performing degrading and dangerous sexual activity'.[167]

2.130 Dr Michael Salter also stated that ongoing trauma can result in neurological and psychological development being interrupted, which can result in increased risk for autoimmune disease and serious psychiatric conditions later in life.[168]

Psychological and physical safety

2.131 At the most extreme end of harms, online harms can transverse into the offline world and pose a significant threat to a victim's psychological and physical safety.

2.132 Victims of online harm can feel extreme fear and threat. Witnesses to the inquiry described feeling a sense of fear and terror as a result of being abused online.[169] The impact of harms such as technology-facilitated abuse can result in victims feeling a sense of 'exhaustion, despair and hopelessness'. [170] Targets of online abuse were also said to be reluctant to participate in online spaces due to fear of being attacked, and for some pulling out of social media engagement altogether to the detriment of their personal lives and careers.[171]

2.133 The sense of fear and threat can go further into fears for personal and physical safety. 'Doxxing', where a person's private information (such as address or contact information) is released online maliciously, can result in threatening conduct such as stalking and harassment.[172] One study found that participants who had been abused had been:

> verbally and physically abused on the street following online harassment, having people come to the houses of women after they were doxxed, only to

---

167 Ms Sonya Ryan, CRF, *Committee Hansard*, 21 December 2021, p. 2.

168 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 15.

169 WESNET, *Submission 25*, p. 5.

170 WESNET, *Submission 25*, p. 5.

171 Ms Nyadol Nyuon, Harmony Alliance, *Committee Hansard*, 22 December 2021, p. 8; Ms Erin Molan, *Committee Hansard*, 18 January 2022, p. 1.

172 Ms Nicole Shackleton, *Submission 28*, p. 7.

> have their child open the front door, or having their animals killed on their private property.[173]

2.134 Witnesses also pointed to the example of the British Member of Parliament, Ms Jo Cox, who was murdered after she was abused, harassed and threatened online.[174]

Impact of COVID-19 on online safety

2.135 The COVID-19 pandemic necessitated a seismic shift for many Australians to move towards online spaces for work, education and communication with friends and family as a result of social restrictions. Some commentators have argued that the increased use of digital products resulted in online risks increasing.

2.136 Ms Kate Everett, founder of Dolly's Dream, stated that the lockdowns resulting from the pandemic resulted in younger people having increased screen time, which exposes children at extremely young ages to online platforms.[175] The Centre for Excellence in Child and Family Welfare also reported that older children were using online platforms significantly more as a result of social restriction measures, particularly in states such as Victoria.[176]

2.137 However, Professor Amanda Third stated that the concerns regarding proliferating online abuse during the pandemic may not be founded:

> What we've seen in the context of the pandemic is an increased use of digital technology and really some of the reconfiguration of children's digital media practices over this two-year period. There is some hypothesis that, as a consequence of that intensified use of digital technology, there has been a heightened exposure of children to forms of online harm. Particularly, there are concerns about cyberbullying, about child sexual exploitation and so on. However, the evidence that we have to date is not yet concrete. Many of those impacts are not yet well documented by rigorous research. We have some strong indications that children may have been exposed to more intense forms of harm, but, at the same time too, we don't really know the full extent of that increased exposure and also we don't know whether the harms that

---

173 Ms Nicole Shackleton, *Submission 28*, p. 7.

174 Ms Nicole Shackleton, *Submission 28*, p. 7.

175 Ms Kate Everett, Founder, Dolly's Dream, *Committee Hansard*, 27 January 2022, p. 3.

176 Ms Deborah Tsorbaris, Chief Executive Officer, Centre for Excellence in Child and Family Welfare, *Committee Hansard*, 27 January 2022, pp 3-4.

> potentially arise are going to extend into the future as we emerge beyond the pandemic. It is quite possible that, as we move out of the pandemic, we will recalibrate and that things will in a sense become more balanced.[177]

2.138 Further, while Dr Third acknowledged that the pandemic had been damaging to children's mental health, there was not a clear causal link between the pandemic and children's mental health. She explained:

> [T]he challenge that's there for us is that the pressure—because we don't know enough yet—that's on children's mental health today is the output, if you like, of a very complex set of dynamics. It is very, very difficult in this scenario to point to causal connections. For example, it's very tempting to say, 'Children have spent a lot more time using social media platforms and other forms of technology to connect and to maintain their education and so on, and this could be a cause of the pressure that they're experiencing,' but of course there are also other shifts and changes happening in children's lives at the same time. Their physical activity has been inevitably reduced because they haven't had capacity to go outside and exercise in the ways they might have. They have had very little time face to face with peers and they've had huge disruptions to their routines and so on. As we unpack this question, there is a lot to be cognisant of and a lot of dynamics to hold in the balance.[178]

2.139 It was also noted by the Isolated Children's Parents' Association that for many children in regional or remote areas, the challenges posed by pandemic-related isolation were not new:

> I think that we really need to remember that geographically isolated children have been doing this for a very long time. They have to have access to online platforms for their education, for starters. And then when a geographically isolated family sends their child to boarding school, they don't want them to be isolated at the boarding school; they wanted them to be able to have contact with their family and with the outside world while they are there. I think while safeguards need to be put in place, we also need to remember that these tools are powerful tools and very useful tools in some ways, so we need to be careful to ensure that the range of availabilities to them are used in a way that is effective and efficient in the unique circumstances that they find themselves in, because completely removing their access or limiting their access too much

---

177 Professor Amanda Third, Young and Resilient Research Centre, *Committee Hansard*, 21 December 2021, p. 26.

178 Professor Amanda Third, Young and Resilient Research Centre, *Committee Hansard*, 21 December 2021, p. 29.

> may lead to the opposite problem, whereby we have underhand things happening because we don't know about it.[179]

Committee comment

2.140 Positive outcomes can result for Australians of all ages and personal backgrounds from online use. The internet has produced, and will continue to develop, amazing and previously unthinkable uses to enhance the lives of people from all walks of life.

2.141 Having said that, the online abuse described by witnesses is unacceptable in modern society. The Committee does not find it in any way tolerable that the most vulnerable groups in society – including children and young people, women, migrants and refugees, people living with disability and others – often experience harm more than any other group. This situation further marginalises these groups by driving them out of the public square and denying them their rights to participate in public discourse as equal citizens.

2.142 Following this point, online culture does not exist in isolation from the offline world. This is reflective of the fact that the most vulnerable groups in offline society are the most likely to be the targets of abuse in online settings as well. This indicates that cultural change may not be possible online until it is addressed in the offline world as well. Broader change in society to ensure the fair, equitable and safe treatment of all Australians, particularly the most vulnerable, is necessary in improving online safety.

2.143 Furthermore, it is important for all Australians to accept that they have a role to play in improving online safety. While technology companies and government have critical roles to play in managing online safety, the Committee is of the view that behind every harmful action online is a person who has chosen to behave this way. It is not until we fundamentally change what we believe is acceptable online conduct that we can truly address online harm.

2.144 The Committee believes that this message should be utilised in an educational campaign directed at all Australians, focusing on digital citizenship. By encouraging Australians to consider the nature in which they engage and act as responsible citizens in the online world, broader cultural change may be possible, both in the digital space and offline.

---

179 Ms Alana Moller, President, Isolated Children's Parents' Association, *Committee Hansard*, 27 January 2022, p. 6.

Recommendation 3

2.145 The Committee recommends that the eSafety Commissioner undertakes research focusing on how broader cultural change can be achieved in online settings.

Recommendation 4

2.146 Subject to the findings in Recommendation 3, the Committee recommends that the Australian Government establishes an educational and awareness campaign targeted at all Australians, focusing on digital citizenship, civics and respectful online interaction.

Recommendation 5

2.147 The Committee recommends that the eSafety Commissioner examine the extent to which social media companies actively prevent:

- **recidivism of bad actors,**
- **pile-ons or volumetric attacks, and**
- **harms across multiple platforms.**

2.148 The eSafety Commissioner should then provide the Australian Government with options for a regulatory framework, including penalties for repeated failures.

2.149 The Committee is mindful of the significant levels of harms caused to Australians who experience online abuse or other forms of significant harm. The Committee was particularly concerned regarding the most serious harms, such as child exploitation, image-based abuse and technology-facilitated abuse, which significantly impact and traumatise victims. It also noted the evidence of Dr Michael Salter who explained that services in relation to recovery from complex trauma in particular are limited in an Australian context.[180]

2.150 Online harm leaves a long trail of trauma on its victims and creates a fundamental sense of not feeling safe, both online and offline. This feeling is accentuated by the inescapability of online harm – while harm experienced

[180] Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 14.

in the schoolyard, workplace or home is unquestionably traumatic, the evidence received by the Committee indicated that harm experienced online can impact every aspect of a person's life. Victims feel a sense that their abuser is always with them, regardless of the physical distance between them. Further, the broad audience that online abuse can attract can accentuate the harm experienced.

2.151 It is appropriate and necessary for survivors of online abuse in all forms to access not only a sense of resolution to their experiences, but also treatment and recovery from ongoing trauma. The Committee encourages research bodies to examine ways in how best to treat people who have experienced online abuse. It is also of the view that the healthcare sector should work with bodies such as the eSafety Commissioner, the Department of Infrastructure, Transport, Regional Development and Communications, and the Department of Health in addressing the need for trauma recovery services.

2.152 The Committee's evidence indicated that many victims attempt to report their experience of abuse to multiple agencies, including police and eSafety. This can potentially retraumatise victims, not only by having to retell and relive their experiences many times over, but also by repeatedly being told that agencies cannot assist in the way that a victim might expect.

2.153 The Committee is of the view that this experience may not always be trauma-informed and could potentially risk victims' mental health. A single point of entry into a complaints' reporting scheme, such as through the eSafety Commissioner for example, could reduce the traumatic load on victims when seeking help, and would encourage coordination and awareness between Australian Government agencies and state- or territory-based police forces.

Recommendation 6

2.154 The Committee recommends that the Office of the eSafety Commissioner be provided with adequate appropriations to establish and manage an online single point of entry service for victims of online abuse to report complaints and be directed to the most appropriate reporting venue, dependent on whether their complaints meet the requisite threshold, and in consideration of a variety of audiences such as children, parents/carers, women, people from culturally and linguistically diverse backgrounds, and other relevant vulnerable groups.

2.155 The Committee took particular note of the evidence received in relation to technology-facilitated abuse, which was highlighted by multiple stakeholders – including eSafety – as a critical issue in the online safety environment. The forms of harm that were identified and the all-encompassing effects on the victims of this form of abuse represent a dangerous development in family violence.

2.156 Further work in this space is required to understand how technology-facilitated abuse manifests, the ways in which technology is utilised to coerce and harm others in family violence situations, and the methods in which digital platforms and government agencies can reduce the prevalence of such behaviour.

2.157 The best forum for this work would be a future inquiry conducted by the proposed House Standing Committee in relation to online matters. This inquiry would seek advice from a broad range of stakeholders, including digital platforms, banks, and women's safety groups, to identify an appropriate course of regulatory reform.

2.158 However, it was clear from the evidence provided to the Committee through submissions and public hearings that this challenge of technology-facilitated abuse is of such a concern and scale that more support is required to support victims through existing Australian Government programs.

2.159 As such, the Committee strongly supports a significant increase in funding for such support, including specialised counselling and support services for victims of technology-facilitated abuse in the next National Action Plan to End Violence Against Women and Children 2022-2032.

Recommendation 7

2.160 The Committee recommends that the Australian Government refer to the proposed House Standing Committee on Internet, Online Safety and Technological Matters, or another committee with relevant focus and expertise, an inquiry into technology-facilitated abuse, with terms of reference including:

- **The nature and prevalence of technology-facilitated abuse;**
- **Responses from digital platforms and online entities in addressing technology-facilitated abuse, including how platforms can increase the safety of their users; and**

- **How technology-facilitated abuse is regulated at law, including potential models for reform.**

Recommendation 8

2.161 The Committee recommends that the Australian Government significantly increase funding to support victims of technology-facilitated abuse, through existing Australian Government-funded programs. This should include additional funding for specialised counselling and support services for victims; and be incorporated in the next National Action Plan to End Violence Against Women and Children 2022-2032.

2.162 Finally, the Committee reiterates the thanks expressed in Chapter 1 to witnesses who provided their experiences of online harm. The Committee could not have developed its understanding of the prevalence, nature and impact of online harm, or have formulated its response to the evidence and its recommendations, without these powerful stories. The Committee commends the courage and resilience displayed by these witnesses.

3. I'm Concerned About This Post

Technological Features that Facilitate Online Harm

Overview

3.1 The question of the role of technology itself in causing harm is divisive amongst industry participants and users. One view argues that technology is by its nature neutral and only a conduit for those who use it to harm others. An alternative view suggests that certain features of digital platforms can further amplify harmful content that is caused to the point that they become a source of harm themselves.

3.2 Social media and other digital companies utilise a number of technological tools and devices in order to provide users with their online experience. In recent years, many of these have been identified as contributing to online harm and causing it to proliferate. Major technology companies working in online spaces have responded to concerns regarding online safety in varying ways, and with varying degrees of success.

3.3 Social media and digital companies more broadly have responded to concerns regarding online safety on their platforms in a variety of different ways, depending on the nature and functioning of their service. Despite this, there is widespread criticism that these measures are not sufficient to adequately address harm.

3.4 This chapter examines how social media and digital companies have responded to the challenges posed by online safety, and the existing

technological tools utilised by these entities that pose concerns for those involved with online safety matters.

Overview of social media companies' online safety responses

3.5 This section provides an overview of measures that major social media and other key digital platforms have taken in response to online safety concerns.

Meta (including Facebook, Messenger, Instagram and WhatsApp)

3.6 Meta Platforms (Meta) is the parent company of multiple social media platforms such as Facebook, Instagram and WhatsApp. It was established in 2004 and is considered one of the Big Five American technology firms.

3.7 Meta's general approach to safety in relation to its social media products hinges on its policies, known as the Community Standards, which outline what is and is not permitted on the platform's services. Meta stated that these Community Standards are based on a set of values prioritising online safety and combating abuse, alongside values of privacy, authenticity, voice and dignity.[1] The Community Standards prohibit certain content themes, such as hate speech, child exploitation, suicide and/or self-injury, violent or abhorrent content, and bullying and harassment.[2] Updates are made regularly to the Community Standards based on current events online and offline.[3]

3.8 Some of the features that Meta have introduced to address online safety include:

- Basic tools for users to manage their experiences of the platforms, including Block, Report, Hide, Restrict and Unfollow functions, in addition to other user-control tools such as managing comments to delete and restrict unwanted interactions, tagging controls, and controls over who can send direct or private messages to them;[4]

1 Meta, *Submission 49*, p. 7.

2 Meta, *Submission 49*, p. 7.

3 Meta, *Submission 49*, p. 8.

4 Meta, *Submission 49*, p. 10.

- Issuing warnings or discouragement to users if they draft a comment or message which resembles a bullying or harassment comment;[5]
- Establishing an Oversight Board, populated by 40 experts in human rights and technology, to make binding rulings on 'difficult and significant decisions about content' in relation to Facebook and Instagram content;[6]
- Guidance towards authoritative sources of information when users search for particular topics, such as 'domestic violence', or blurring triggering images if a user searches for self-harm or eating disorders;[7] and
- Resources and learning modules focusing on online safety and the tools Meta offers to manage users' experiences.[8]

3.9 Meta also has a Safety Advisory Board for its global operations, which consists of a number of organisations and individuals who are experts in online safety and which provides advice to inform Meta's safety policies.[9]

3.10 In addressing concerns regarding young people using its platforms, Meta has worked with the eSafety Commissioner in addition to other international governments and industry partners to create a Youth Design Guide, which suggests the following principles be incorporated into any Meta product aimed at young people: '(1) designing for different levels of maturity; (2) empowering young people with meaningful transparency and control; and (3) undertaking data education for young people'.[10]

3.11 Meta has also acknowledged and taken action on other vulnerable user groups. For example, having identified women as a vulnerable user group on its services, Meta has implemented the following policies and features:

- Ensuring gendered and culturally specific forms of harm are catered for in policies regarding prohibited content (e.g. expanding the bullying and harassment policy to incorporate stricter regulation of female-gendered cursing);

---

5 Meta, *Submission 49*, p. 3.

6 Meta, *Submission 49*, p. 3.

7 Meta, *Submission 49*, pp 11-12.

8 Meta, *Submission 49*, p. 12.

9 Meta, *Submission 49*, p. 2.

10 Meta, *Submission 49*, p. 15.

- Investing and leading in combatting non-consensual sharing of images (including leading StopNCII.org, which enables people concerned that their image has been shared to work with online platforms to stop the proliferation of the content); and
- Working with groups such as the Women's Services Network (WESNET) and 1800 RESPECT to promote messaging regarding family violence and link users to these services.[11]

Google (including Search and YouTube)

3.12 Google is an American technology company which produces services and products primarily in relation to the online market. Founded in 1998, the company operates the search engine Google Search, in addition to other Google products, and also administers YouTube.

3.13 Digital companies sitting outside the social media space have different online safety considerations, depending on the platform they administer and its functionalities. As primarily a search engine in addition to other platforms, Google Search has implemented a number of online safety policies to address harm on its service. Some of these policies include:

- The Safe Search tool, which filters out explicit content in Google search results across images, videos and websites, and is a default setting for all signed-in users under the age of 13 with accounts managed by Family Link (a parent-established account which stipulates digital ground rules);
- Removal of images of those under 18 years of age by request from the minor or their parent or guardian;
- Removal of non-consensual explicit images by request; and
- Content policies for particular features (such as autocomplete) to prevent dangerous or violent content from appearing in searches.[12]

3.14 In administering YouTube, Google Australia has focused on building products with built-in safety features specifically for children. Its two main products aimed at children are YouTube Kids, which is a specific child-centred app which uses filters and content moderation to provide safe content, and the Supervised Experience, which is available for parents using Family Link to enable children to access the main YouTube service but with adjustments to support children's presence (including limiting types of

11 Meta, *Submission 49*, pp 21-26.

12 Google Australia, *Submission 30*, pp 3-4.

advertisements, disabling video uploading, livestreams and reading or writing comments).[13]

3.15 Other safety policies applied by Google across the entire platform include:

- Prohibition of 'sexually gratifying' content;
- Prohibition of content that 'endangers the emotional and physical well-being of minors', which includes the sexualisation of minors, harmful or dangerous acts of minors, inflicting emotional distress, misleading content which is directed at minors but contains inappropriate themes, and cyberbullying;[14]
- Prohibition of content encouraging dangerous or illegal activities, such as dangerous challenges, instructional videos on hurting yourself or others, and content praising eating disorders; and
- Placing content warnings on particular kinds of content, such as content relating to topics such as suicide or self-harm.[15]

3.16 Google Australia also plans to disable certain functions for young people, including Location History.[16]

Twitter

3.17 Founded in 2006, Twitter is an American-based company, which runs a micro-blogging and social media service of the same name.

3.18 Twitter has a set of Rules and Terms of Service, which set out appropriate behaviour. Twitter describes these policies as 'living documents':

> We're updating them every week and every month, given how rapid the changes are around these debates and around how we can move forward to make sure that women feel safe on the platform and that all vulnerable and underrepresented groups have a place and a voice until safe and welcome on Twitter.[17]

3.19 A non-exhaustive list of Twitter's online safety features include:

---

13 Google Australia, *Submission 30*, pp 4-5.

14 YouTube, *Child safety policy*, available at: https://support.google.com/youtube/answer/2801999 (accessed 4 February 2022).

15 Google Australia, *Submission 30*, p. 5.

16 Google Australia, *Submission 30*, p. 8.

17 Ms Kathleen Reen, Senior Director of Public Policy, Asia-Pacific, Twitter, *Committee Hansard*, 21 January 2022, p. 17.

- Basic functions for users to control their experience, including blocking other users, reporting functions and privacy settings to prevent direct messages or Tweets from unknown users;
- Changes to control over algorithms, including the ability to turn off the default ranking system;[18]
- Policies in relation to particular forms of harm, including policies on hateful conduct;[19]
- A Trust & Safety Council which develops products and features in addition to improving Twitter's rules;[20]
- A Tips function that makes recommendations to users to improve online safety and privacy;
- The Safety Mode, which blocks a person from using their account for seven days in cases of Rule or Terms of Service breaches;
- Education resources, including for parents, young people and vulnerable groups.[21]

3.20 In addition, Twitter utilises a number of accountability features, including:

- A Responsible Machine Learning Initiative, which provides information in relation to the operation of its algorithms and how Twitter has been attempting to improve it;[22]
- The publication of biannual Transparency Reports, containing information about Twitter's enforcement of its Rules;
- The creation of the Twitter Transparency Centre, which covers a broad range of topics such as information requests, removal requests, Rules enforcement and other matters; and
- Features aimed at academics in order to facilitate open access and developments across a wide network of experts to improve online safety and technology. [23]

Snapchat

3.21 Snap Inc. (Snap) is the parent company of the social media and camera-based app Snapchat. Established in 2011, Snapchat is an instant messaging

18 Twitter, *Submission 50*, p. 5.

19 Twitter, *Submission 50*, p. 8.

20 Twitter, *Submission 50*, p. 9.

21 Twitter, *Submission 50*, p. 10.

22 Twitter, *Submission 50*, p. 5.

23 Twitter, *Submission 50*, pp 12-13.

service which differs markedly in its operation in comparison with other social media platforms such as Facebook and Twitter. This includes the lack of an open and uncontrolled News Feed, and a limited number of public spaces on the app, most of which are curated and pre-moderated by the platform. This, Snap argues, prevents the spread of harmful content to large audiences, and avoids the need to utilise artificial intelligence or automated moderation technology to detect harmful content.[24]

3.22 Snap stated that, in developing the Snapchat app, it has followed safety by design and safety by privacy principles from the design stage through to the operations phase of work.[25] It noted that SnapChat is 'designed for private communications (either 1:1 or in limited-size groups), with the aim of encouraging users to interact creatively with their real friends, not strangers'.[26]

3.23 Other safety features on Snap include:

- Community Guidelines which prohibit certain kinds of content and outlining enforcement actions where breaches are identified;
- Reporting tools for harmful content, monitored and actioned by a global Trust & Safety team;
- A default deletion setting which provides that messages and Snaps are deleted from Snap's servers once opened, and Stories on the platform are deleted after 24 hours;
- Utilising technological applications to detect harmful content, including PhotoDNA and CSAI Match in relation to child exploitation material; and
- Privacy features such as not displaying users' friends lists to others and location sharing settings set to off by default.[27]

3.24 Snap also stated that it focuses strongly on the prevention of harm before it occurs. It explained that its focus on limiting content from public broadcast without pre-moderation, and preventing contact from strangers, enables the app to limit harm.[28]

---

24 Snap Inc., *Submission 16*, p. 1.

25 Snap Inc., *Submission 16*, p. 1.

26 Snap Inc., *Submission 16*, p. 2.

27 Snap Inc., *Submission 16*, pp 1-5.

28 Snap Inc., *Submission 16*, p. 2.

TikTok

3.25 TikTok is a self-described 'entertainment' platform which focuses primarily on short videos uploaded by users. Originally Chinese-based as the app Douyin, the international version of the app (now TikTok) was released in 2017, and officially launched in Australia in 2019.

3.26 Similarly to Meta and Twitter, TikTok has emphasised the provision of user control over their experiences of the platform while also working to promote safety online.[29]

3.27 Safety features utilised by TikTok include:

- Terms of Services and Community Guidelines, which set out the prohibited behaviours and content;
- In-app and off-app mechanisms to report harmful content;
- A mix of technological and human-initiated detection and enforcement of harmful content;
- Providing information and guidance in relation to tailoring online experiences via its Australian Safety Centre;
- Automated detection of inappropriate or unkind comments which will prompt users to reconsider posting the comment;
- Parental control features, including a family pairing feature which enables parents and guardians to link their account to their child's account and utilise certain content and privacy settings;
- Youth-specific policies, including turning off direct messages for accounts owned by users between the ages of 13 and 15 years old, default privacy settings, and age restrictions on video sharing, Duet and Stitch functionalities.[30]

How does social media technology create harm?

3.28 Social media platforms are among the most prominent digital actors. The vast majority of Australians engage in social media in order to communicate, engage in business-related activities, and other activities. While there are doubtlessly positive functionalities embedded in social media platforms, concerns have been raised that certain elements of social media technology have the potential to cause harm and intensify existing harm.

---

29 TikTok Australia, *Submission 57*, p. 1.

30 TikTok Australia, *Submission 57*, pp 2-5.

Social media systems' design

3.29 A common and strongly held sentiment heard from many witnesses was that social media services and platforms have almost universally not been designed with users' safety and protection in mind. Rather, witnesses argued that social media platforms are primarily designed from a profitability perspective, which overrides the need to provide user safety.

3.30 Dr Hany Farid explained how the social media industry is reliant on extreme content to foster profitability:

> You have to understand that social media—and I agree this is not entirely a social media problem but let me focus on that for a minute—give away their product for free. They are in the ad-delivery, engagement-driving business, which means engagement in and of itself is the product. As it turns out, humans are sort of awful, so the most hateful, salacious, outrageous, conspiratorial conduct is what engages. It's not that they're not able to do this; it's against their financial interests.[31]

3.31 This perspective was similarly expressed by Mr Peter Lewis, Director, Centre for Responsible Technology, who argued that the industry's business model was essentially 'not just providing the service but observing everything we do and then making money out of those actions'.[32] He put the view that the industry is reliant on users spending as much time on its platforms as possible. Mr Lewis stated that social media companies were conscious that 'if you wanted to totally maximise profits, you make your engagement as intense an experience as you can', and that this behaviour was not necessarily in the public interest.[33]

3.32 Dr Michael Salter argued that the social media industry was designed primarily as a profit-generating business, rather than as a tool regulating the content provided to young people, which is reflected in its platform:

> Social media has been designed and marketed to be particularly attractive to children and young people, but it was built without regard for user safety. The underlying business model aims to maximise profit by maximising the frictionless circulation of content and contact between users, with a minimum

---

31 Dr Hany Farid, *Committee Hansard*, 28 January 2022, p. 5.

32 Mr Peter Lewis, Director, Centre for Responsible Technology, *Committee Hansard*, 28 January 2022, p. 13.

33 Mr Peter Lewis, Centre for Responsible Technology, *Committee Hansard*, 28 January 2022, p. 12.

> of expenditure on content moderation and oversight. This model has simply proved, over the last 25 years, to be incompatible with child protection.[34]

3.33 The Centre for Digital Wellbeing (CDW) put the view that social media platforms are driven by their profitability-based business models to 'target and manipulate our social characteristics'.[35] They further explained how this occurs and its impact:

> The harms of this model are produced both through the algorithms they use to drive engagement and derive profits and through their design features such as filters, shares, likes and infinite scrolling. Such platforms engage us by hyperstimulating us and artificially producing validation. They feed belonging by generating collective outrage, and they cultivate and manipulate identity through algorithmically curated content.[36]

3.34 Ms Frances Haugen, a former Facebook (now Meta) employee who then became a whistleblower, explained that Facebook's choice to use algorithms that promote extreme content is an example of the company's prioritisation of profitability at the expense of safety.[37] Another example was pointed to by Dr Hany Farid, who explained that Facebook's banning of adult pornography was an instance where profitability overrode considerations of safety:

> In the very early days of Facebook and YouTube, they banned legal adult pornography. When Mark Zuckerberg and Jack Dorsey tell you how much they love the First Amendment, would you please ask them why they banned perfectly protected speech. The reason they did it was that it was bad for business because advertisers don't want their ads running against sexually explicit material. So what did they do? They developed very good technology that, for the most part, keeps legal speech off Facebook and YouTube. It was not a technological problem.[38]

3.35 Victims of online harm also expressed frustration that reporting online harm would not lead to change in the platforms due to their business models

---

34 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 9.

35 Ms Carla Wilshire, Chair, Centre for Digital Wellbeing (CDW), *Committee Hansard*, 21 January 2022, p. 28.

36 Ms Carla Wilshire, CDW, *Committee Hansard*, 21 January 2022, p. 28.

37 Ms Frances Haugen, *Committee Hansard*, 3 February 2022, p. 7.

38 Dr Hany Farid, *Committee Hansard*, 28 January 2022, p. 5.

being based on advertising and giving content as broad an audience as possible. Ms Erin Molan stated:

> Even reporting things on Instagram, you feel like you're banging your head against a brick wall, because you look at their business model and […] advertising is the biggest thing for them. The more people they have on their account, the more advertising they get. They don't want to verify every single user on their account. They'd love one person to have 8,000 accounts because it gives them more people to sell to advertisers. So it just feels like you're banging your head against a brick wall. I can't see them ever doing anything off their own bat or ever cleaning it up themselves.[39]

Social media technology not being used for individual or group harm

3.36 The inquiry demonstrated that there appears to be a double-standard in how social media companies treat different categories of harm. Evidence to the Committee suggested that while social media companies have the technology and capabilities to appropriately address online harm when addressed at individuals or groups, this does not lead to those harms being adequately addressed.

3.37 As outlined above, many major social media companies have policies on their acceptable standards of behaviour in addition to methods of detection. It was put, however, that the detection and removal of particular forms harmful content and behaviours was not being utilised effectively.

3.38 The Committee observed during its examination of major social media companies such as Meta and Twitter that while these platforms were proactive and responsive in the detection and removal of certain forms of harmful content, such as misinformation and disinformation in relation to medical and electoral information, their responses to harm directed at particular individuals or groups were not as strong.

3.39 It was put to Google, for example, that while they heavily moderated content in relation to COVID-19 treatments as a form of misinformation and disinformation, there were examples of content containing abuse directed at public officials on YouTube which the platform had not removed when reported. Google responded by stating that, when reviewing complaints, its

---

39 Ms Erin Molan, *Committee Hansard*, 18 January 2022, p. 4.

safety teams considered a range of contextual information, including whether the recipient of the abuse in question is a public official.[40]

3.40 Other social media services such as Meta were accused of not focusing its energies on sources of individual and community-based harm, and focusing on particular topics such as misinformation. In response to claims that Meta had removed the account of a public official due to misinformation and disinformation concerns, Meta stated that it had acted to enforce its policies which had been informed by expert and leading advice.[41]

3.41 Social media companies, often multinational corporations tending to numerous jurisdictions around the world, were also argued to be applying a 'one size fits all' approach rather than taking consideration of local laws and standards.[42]

3.42 A more holistic approach to consideration of online harm prevention was well articulated by Dr Kate Hall, Head of Mental Health at the Australian Football League (AFL), who said:

> I think what we're understanding about human behaviour is that, particularly for young people, or when people begin this type of behaviour, the peer and social norms of, I guess, guardrails are critical on making lasting change. Whilst a one-off deterrent might bring it to attention, those many other policy pieces and protectors in place then step in for something more sustainable, as an intervention in and of itself to deter others from this… I want to reiterate what Tanya said around unmasking and particularly the anonymity. I do think that's a very critical piece of this puzzle about why this behaviour that is so harmful to others is able to flourish and to grow. It doesn't have to be a very stringent act. It can actually start earlier, when people begin to engage in things that are perceived as harmful to others. We would want them held accountable early, instead of waiting for the behaviour to escalate to that point. All anti-social behaviour, I believe, at some point starts with a test. Then when there's no action, it grows and becomes more and more harmful.[43]

Lack of a duty of care on social media platforms

---

40 Ms Lucinda Longcroft, Director, Government Affairs and Public Policy, Google Australia and New Zealand, *Committee Hansard,* 20 January 2022, p. 11.

41 Ms Mia Garlick, Regional Director for Policy, Australia, New Zealand and the Pacific Islands, Meta, *Committee Hansard*, 20 January 2022, pp 23-24.

42 Scarlet Alliance, Australian Sex Workers Association, *Submission 85*, p. 7.

43 Dr Kate Hall, Head of Mental Health and Wellbeing, Australian Football League (AFL), *Committee Hansard*, 1 February 2022, p. 15.

3.43 Representatives from the social media and digital industry stated that many companies had implemented policies to protect users, but further action would require cooperation from multiple stakeholders.[44] Ms Sunita Bose, Managing Director, Digital Industry Group Inc. (DIGI), explained that the lack of consistent standards across the digital industry made it difficult to ensure safety across all platforms, which DIGI was working to address through its work in drafting new industry codes of practice.[45]

3.44 As part of its inquiry into the adequacy of existing offences in the Commonwealth Criminal Code and of state and territory criminal laws to capture cyberbullying, the Senate Legal and Constitutional Affairs References Committee recommended that the Australian Government 'legislate to create a duty of care on social media platforms to ensure the safety of their users'.[46] In response, the Australian Government noted this recommendation, indicating that it would monitor developments in other jurisdictions regarding a duty of care at law for digital platforms.[47] The Government's response stated:

> The Government considers that online safety is a shared responsibility, and that content and behaviour which is prohibited offline should also be prohibited online. The Government considers that social media platforms and other technology firms need to recognise that their responsibility for tackling harmful behaviours and content goes hand-in-hand with their influential and important position within Australian society. It is particularly important that industry participants whose products and services are used by children take appropriate action to uphold the safety of their users.[48]

---

44 Ms Sunita Bose, Managing Director, Digital Industry Group Inc. (DIGI), *Committee Hansard*, 20 January 2022, p. 37.

45 Ms Sunita Bose, DIGI, *Committee Hansard*, 20 January 2022, pp 37-38.

46 Senate Legal and Constitutional Affairs References Committee, *Adequacy of existing offences in the Commonwealth Criminal Code and of state and territory criminal laws to capture cyberbullying*, March 2018, p. 63 (Recommendation 8).

47 Australian Government, *Government response to the Senate Legal and Constitutional Affairs References Committee's report for its inquiry into the adequacy of existing offences in the Commonwealth Criminal Code and of state and territory criminal laws to capture cyberbullying* (Government Response), April 2021, available at: https://www.aph.gov.au/DocumentStore.ashx?id=e1e15f74-60db-4894-960b-4e89ddcf9834 (accessed 10 January 2022) p. 19.

48 Australian Government, *Government Response*, April 2021, available at: https://www.aph.gov.au/DocumentStore.ashx?id=e1e15f74-60db-4894-960b-4e89ddcf9834 (accessed 10 January 2022) p. 19.

3.45 The response further stated that, in the event that digital platforms and companies 'fall short of community standards', the Australian Government would consider how best to protect Australians online via regulatory means.[49]

Lack of effective detection practices for harmful content

3.46 Social media platforms generally have policies and systems in place for violations to its terms of service, rules and community standards. Evidence by multiple digital platforms indicated that most social media companies use a combination of common elements to detect and remove harmful content:

- Detecting harmful content prior to its publication and distribution on platforms, utilising artificial technology (AI) and other forms of automated detection systems in combination with human oversight teams; and
- Encouraging users to report content when it is found to prevent its proliferation and avoid causing further harm.

3.47 Most online platforms appeared to predominantly utilise automated and AI systems to detect harmful content at the first instance.

3.48 Effective tools to detect forms of harm, particularly in relation to CSAM and other extremely harmful material were argued by submitters to be critically important for addressing online harm in digital platforms. Nonetheless, some witnesses were critical of efforts being made by digital companies in detecting online harm.

Detecting child exploitation material

3.49 Witnesses drew particular attention and urgency to the issue of the detection of CSAM content, and the role of social media companies and digital companies. Dr Michael Salter, an expert in CSAM and online harm, put the view that social media platforms were being used by predators to contact and groom children and young people, while the platforms appear to not be able to address the issue adequately:

> Social media companies tell us that they are just as concerned about child safety as we are, but the amount of child sexual abuse material reported to

[49] Australian Government, *Government Response*, April 2021, available at: https://www.aph.gov.au/DocumentStore.ashx?id=e1e15f74-60db-4894-960b-4e89ddcf9834 (accessed 10 January 2022) p. 19.

> Australian and overseas authorities increases every year. Prosecutions for child sex exploitation offences in Australia are also increasing year on year. Abusers are using social media to circulate child sexual abuse material. They are using social media to contact and sexually harass children and to manipulate and extort them into producing nude or sexual content. Abusers are also using social media to connect with each other to create online abuse communities to justify their sexual interests and to publicly argue for policy reform that compromises child protection efforts.[50]

3.50 Dr Salter explained that, because social media and digital platforms make representations that they cannot adequately detect CSAM, some users (a number of which were victims of CSAM themselves) were witnessing and reporting CSAM to social media companies and authorities.[51] He explained further:

> Really one of the most distressing examples was this. I spent some time a couple of years ago working with a group of Twitter users who were child abuse survivors. Images had been made of their abuse. And they were on Twitter flagging child sex abuse material on Twitter. I have to say that the content that I saw when I was doing research with this group was the most serious content you can imagine on Twitter. It included videos of infant children being raped. It was absolutely horrific. This was content that was widely circulating on Twitter, and it was up to child sexual abuse survivors themselves to hunt down this content and report it to Twitter because there didn't seem to be any effective proactive measures by Twitter to take that content offline.[52]

3.51 In March 2020, the Five Country Ministerial (consisting of the governments of Australia, Canada, New Zealand, the United Kingdom and the United States) launched the Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse (Voluntary Principles). The Voluntary Principles were designed in partnership with industry actors such as Meta, Google, Snap, TikTok, Twitter, non-government groups, academics and others, to encourage awareness of the issue and prompt action. The Voluntary Principle' themes, which contain 11 principles, include:

- The prevention of child sexual abuse material;
- The targeting of online grooming behaviour;
- The targeting of livestreaming for the purposes of child exploitation;

---

50 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 9.

51 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 11.

52 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 11.

- Preventing the accessibility of child exploitation material in search results and automatic suggestions;
- Adopting appropriate and specialised safety measures for children in particular;
- Responding appropriately to material found, including reporting options for users and alerting authorities; and
- Collaborating and responding to evolving threats.[53]

3.52 Since the adoption of the Voluntary Principles, the Technology Coalition (a global alliance of technology companies, including Meta and others) have announced Project Protect, which establishes the commitment of member companies to address and invest in the prevention of child exploitation, designed to address work over the next 15 years.[54] The Department of Home Affairs (Home Affairs), however, stated that '[i]n almost two years since tech companies endorsed the Voluntary Principles, there is limited evidence as to the degree of implementation and the level of success'.[55]

3.53 Google Australia stated that it was 'committed to stopping the use of our platforms to spread child sexual abuse material (CSAM)'.[56] Google explained that primarily used its Content Safety API and CSAI Match tools in detecting CSAM material, in addition to its Trust & Safety teams to address any incidents as they arise. It stated that it uses 'hashes' to automatically detect certain forms of harmful content prior to it being viewed. Its systems automatically remove content 'only when there is high confidence of a policy violation' and any borderline cases are flagged for review.[57] It stated that this approach resulted in 94 per cent of removed content being flagged by its systems rather than humans, and almost 40 per cent of those videos detected by AI were never viewed.[58] Further, Google Australia stated that it

---

53 Department of Home Affairs (Home Affairs), *Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse*, available at: https://www.homeaffairs.gov.au/news-subsite/files/voluntary-principles-counter-online-child-sexual-exploitation-abuse.pdf (accessed 3 March 2022).

54 Meta, *Submission 49.1*, p. 20.

55 Home Affairs, *Submission 40*, p. 9.

56 Google Australia, *Submission 30*, p. 7.

57 Google Australia, *Submission 30*, p. 17.

58 Google Australia, *Submission 30*, p. 17.

was developing new technology to assist in identifying CSAM content on its services.[59]

3.54 Like Google, Snap Inc. stated that it also utilises technology such as PhotoDNA and CSAI Match which assists in identifying known images and videos of CSAM.[60]

3.55 Meta provides limited publicly available information in relation to how it detects CSAM and other forms of child exploitation material. In Meta's Community Standards Enforcement Report for Q4/2021 period, it stated that approximately 97.5 per cent of violating content in relation to child exploitation material was identified by Meta itself, while the remaining 2.5 per cent of material was reported by users.[61] In evidence provided to the Joint Standing Committee on Law Enforcement, Meta asserts that it has developed two new kinds of detection technology for images and videos, which it has made open source for other platforms to use.[62] This information is extremely difficult to find in the public sphere aside from this submission. Further, Meta does not provide any further details in relation to its detection practices aside from reference to its public commitments to reduce child exploitation material online.[63]

Detecting abusive content

3.56 Concerns were raised by a number of witnesses that the detection of harmful or abusive content more generally on social media platforms was insufficient to appropriately protect users from harm.

3.57 Witnesses to the inquiry reported three commonly experienced issues:

- The failure of social media platforms to adequately uphold their terms of use, including community standards and policies;

---

59 Google Australia, *Submission 30*, p. 8.

60 Snap Inc., *Submission 16*, p. 3.

61 Meta Transparency Centre, *Community Standards Enforcement Report – Child Endangerment: Nudity and Physical Abuse and Sexual Exploitation*, available at: https://transparency.fb.com/data/community-standards-enforcement/child-nudity-and-sexual-exploitation/facebook/ (accessed 3 March 2022).

62 Facebook, *Submission 24*, Joint Standing Committee on Law Enforcement inquiry into law enforcement capabilities in relation to child exploitation, available at: https://www.aph.gov.au/DocumentStore.ashx?id=25ed9734-24c6-429d-8f3d-64b5a6eb7b13&subId=712353 (accessed 3 March 2022), p. 2.

63 Meta, *Submission 49.1*, p. 20.

- The lack of detection by social media platforms of harmful content, including the failure of algorithms to detect all forms of harmful content; and
- The subsequent difficulty in having it removed or seeking redress.

3.58 For example, footballer Ms Tayla Harris explained to the Committee that she has received a significant amount of online abuse which is gender-based, but that is not detected by social media platforms, and the removal of that content often requires substantial and lengthy engagement with platforms individually.[64] The experiences described by Ms Harris were commonly experienced by many witnesses and submitters.

3.59 Twitter was asked about its approach to managing abuse and harassment at scale directed at individuals. It stated that it factored in two considerations when responding to this issue: examining how the platform may be incentivising actors that take advantage of the service and circumvent its policies and rules, and examining how it can reduce victims' role in reporting abuse by utilising technology:

> Under looking at the design of Twitter, we've really strengthened our policies around the platform manipulation … where we have not only added new teams but also got a lot better at using machine learning to identify when we receive some sort of swarm or sudden uptick in attention or abuse, especially targeted at specific or individual accounts. We are then able to then have that surfaced and flagged to our Twitter service or concept moderation team, which is able to look at this against our policies and then take coordinated action across the accounts that are participating in that behaviour instead of trying to do it at that one-off back and forth. We've seen really strong moves towards being able to take this down at scale and we've seen this actually result in fewer abuse reports having to be submitted to our teams to be able to see that this kind of attention needs to be given to a specific case or a specific group of accounts.[65]

3.60 The platform stated that it had also introduced a new feature which prompted users to reconsider posting a Tweet if abusive language or content had been detected, which was reportedly positively accepted by users.[66] Twitter also noted that it has a function which enables people without user

---

64 Ms Tayla Harris, *Committee Hansard*, 1 February 2022, p. 20.

65 Ms Kara Hinesley, Director of Public Policy, Australia and New Zealand, Twitter, *Committee Hansard*, 21 January 2022, pp 11-12.

66 Ms Kara Hinesley, Twitter, *Committee Hansard*, 21 January 2022, p. 16.

accounts to make reports if they become aware of inappropriate content on the platform.[67]

3.61 Twitter stated that, while it did not tolerate abuse and harassment on its platform, it did 'allow for certain inflammatory or strident language' depending on the broader context.[68] In investigating a reported allegation of abuse or harassment using abusive language, Twitter stated that it would examine the conduct and whether it was compliant with its rules and terms and conditions. It would also look at factors such as whether the involved user accounts followed each other and other 'behavioural signals' which would provide a clearer picture of the situation.[69] Twitter explained:

> [T]here are instances where we have seen certain phrases or strident language that are sometimes used in a slang way, as a community or even sometimes as terms of friendship. Our former VP Del Harvey used to use the example that 'hey, b-i-t-c-h' could be seen as a greeting to a friend or could be seen as abuse, depending on the context. As you'll see, especially within a younger demographic, there are terms that are used in a way that might seem on their face to be abusive but at other times would be seen as appropriate language between people who know each other.[70]

3.62 Meta similarly argued that context and intent are important in assessing whether material met the definition of abusive material on its platforms.[71] It outlined the ways in which it identifies abusive material on its systems:

> We use human review and developed AI systems to identify many types of bullying and harassment across our platforms. However, as mentioned above, because bullying and harassment is highly personal by nature, using technology to proactively detect these behaviours can be more challenging than other types of violations. It can sometimes be difficult for our systems to distinguish between a bullying comment and a light-hearted joke without knowing the people involved or the nuance of the situation. That's why we also rely on people to report this behaviour to us so we can identify and remove it.[72]

---

67 Ms Kara Hinesley, Twitter, *Committee Hansard*, 21 January 2022, p. 17.

68 Ms Kara Hinesley, Twitter, *Committee Hansard*, 21 January 2022, p. 17.

69 Ms Kara Hinesley, Twitter, *Committee Hansard*, 21 January 2022, p. 17.

70 Ms Kara Hinesley, Twitter, *Committee Hansard*, 21 January 2022, p. 17.

71 Meta, *Submission 49*, p. 16.

72 Meta, *Submission 49*, p. 16.

3.63 Meta explained that it encourages its users to report any harmful content they identify, which then is assessed and 'action the content consistent with our policies'.[73] Meta also stated that it had recently begun investing in detection technology (such as forms of AI and automated detection technology) to identify and remove harmful content before it is seen and reported to the platform.[74] The company reported that its AI systems were found to be increasing the percentage of material identified proactively, increasing to 59.4 per cent of all bullying and harassment material on Facebook from 25.9 per cent a year previously.[75] The AI systems also proactively detected 83.2 per cent of all bullying and harassment material on Instagram.[76]

3.64 Google outlined its detection tools in its submission, although provided limited detailed information:

> We have robust mechanisms to monitor compliance with our policies and to enforce our policies. We rely on a mix of human and technological intervention: we encourage all users to report content that violates our Community Guidelines; we have established the YouTube Trusted Flagger programme, by which individual users, government agencies and NGOs can notify content that violates our Community Guidelines; and we have developed machine learning classifiers to automatically and quickly identify and remove potentially violative content. Content that is found to violate our Community Guidelines is removed; in addition, enforcement may have repercussions for those who violate our policies and may result in channel or account terminations.[77]

3.65 Snap stated that it mostly uses a team of expert analysts to moderate content online, in addition to technological tools to detect abuse such as CSAM.[78] It stated that, given that it is primarily a platform that facilitates communication between two people or small groups, it attempts to seek a balance between the detection of harmful content and respecting the privacy of users.[79]

---

73 Meta, *Submission 49*, p. 8.

74 Meta, *Submission 49*, p. 8.

75 Meta, *Submission 49*, p. 16.

76 Meta, *Submission 49*, p. 16.

77 Google Australia, *Submission 30*, p. 6.

78 Snap Inc., *Submission 16*, p. 3.

79 Snap Inc., *Submission 16*, p. 3.

Attacks on high-profile individuals

3.66 As outlined in Chapter 2, the Committee heard evidence from multiple high-profile witnesses, ranging from news journalists to disability advocates to professional football players, who had experienced online abuse via social media platforms. There is a vast range of online harm that is directed towards individuals who by virtue of their job, position, gender, race or other identification and how these identifications are used to subject people to severe and sometimes sustained online abuse.

3.67 Twitter stated that it was aware of trends that suggested heightened abuse and harassment directed at women in prominent positions, such as politicians and journalists.[80]

3.68 Some of the examples given of online harm directed at public figures include:

> When I first started on *The Footy Show* I noticed some of the commentary—and I would never seek out the commentary written online; I learnt that lesson very early on. Things that were sent directly to me on platforms that I used professionally were just horrific. They were not things like, 'We don't like watching you.' They were things like: 'We will ensure you die. I will hit you with a bus.' These things were so horrific. What they said they would do to me if they ever saw me made me fear for my safety essentially and made me nervous about going outside. There was the detail. People would send me things that they would hope to do to either me or my child.[81]

> Due to me speaking up in defence of my own company and my own sex-based rights, I have received death threats, rape threats, general threats of violence and countess instances of misogynistic abuse.[82]

3.69 When these issues were raised with various platforms, most made a distinction between online harm directed at individuals who had no public profile versus online harm directed at figures who had a high public profile due to their job. Most platforms commented that they had a higher threshold for the takedown of abusive content for public figures, citing freedom of speech as an explanation for this increased threshold. This was evident during the committee hearings with Google and Twitter:

---

80 Ms Kara Hinesley, Twitter, *Committee Hansard*, 21 January 2022, p. 18.

81 Ms Erin Molan, *Committee Hansard*, 18 January 2022, p. 2.

82 Giggle, *Submission 43*, p. 2.

> Chair: Would this comment breach your community standards: would a user of your platform be censored or banned for calling a man or woman a 'whiny little b-i-t-c-h'? I would like a simple yes or no answer at this stage.
>
> Ms Longcroft: I understand that you are referring to a particular case that relates to a public figure here in Australia. Again, those standards would have to take into account the context and the nature of the person who had made the comment. In determining any particular case—and I wouldn't comment on a particular case—those very clear policies would be met.[83]
>
> In terms of any sort of specific phrases or terms, from a Twitter perspective, we do allow for certain inflammatory or strident language. That being said, the statement that you just read out, if it were in the context of targeting an individual or of crossing that threshold into abuse, we would of course review it, under the Twitter rules and terms of service, and take into account any account context. This would depend, again, on who was being @ mentioned, if these accounts followed each other and a number of behavioural signals that would allow us to understand fully what's going on in that situation.[84]

3.70 The Committee considers that there are two challenges with this approach by platforms.

Dehumanisation of public figures in public discourse

3.71 The first is that the different threshold for taking down abusive content online suggests is it acceptable to dehumanise public figures. Experiences of dehumanisation were cited by Ms Erin Molan, Ms Tayla Harris and the Committee Chair.

> You never think that it will impact in the way that it does. As I said, I'm not just talking about people in the public eye, personalities and reality stars. They should all be afforded the same protection as everyone else.[85]
>
> Chair: I have given a very direct example that didn't include the word b-i-t-c-h: 'cavorting whore'. Is there any context that you can think of in Twitter's hateful content policies under which that would be seen to be acceptable and

---

83 Ms Lucinda Longcroft, Director, Government Affairs and Public Policy, Google Australia and New Zealand, *Committee Hansard,* 20 January 2022, p. 7.

84 Ms Kara Hinesley, Director of Public Policy, Australia and New Zealand, Twitter, *Committee Hansard*, 21 January 2022, p. 13.

85 Ms Erin Molan, *Committee Hansard*, 18 January 2022, p. 2.

> not hateful content or abusive content? I will be honest; I can't think of one, but I would like to think of one if you know of one.[86]
>
> 'Every single piece of evidence means this is the exact example of where someone thinks it's acceptable to make awful comments to a stranger online. They get away with that, and what are they going to get away with next? They test things.' That was something that brought the severity of the whole situation to life.[87]

An increase in online abuse towards women in public positions

3.72 The Committee is concerned that allowing a higher threshold for public figures could contribute to the increase in online abuse, particularly against female public figures. Twitter explained this growing tendency in online discourse:

> Unfortunately, it is very common for women in public life to receive those types of threats. We're regularly talking to female journalists as well who are receiving those types of comments.[88]

3.73 Female public figures who presented evidence to the Committee noted that:

> Giggle's App Store and Google Play pages frequently receive 'one star reviews' full of misogynistic slurs, abuse and blatant lies. The goal of the posters is to discourage women from using an online female space and/or to have the female space removed from the market completely… In addition to the abuse towards Giggle, I personally receive online abuse due to my status as the founder and CEO of Giggle and for taking a public stance to protect female-only (single sex) spaces.[89]
>
> It is an increasingly acknowledged reality that social media platforms host a huge amount of vile, threatening, violent, or sexually explicit and pornographic content. Much of this content is directed at women and is posted anonymously. In many cases this explicit and threatening content is publicly visible to children.[90]

---

86 Committee Hansard, 21 January 2022, p. 14.

87 Ms Tayla Harris, *Committee Hansard*, 1 February 2022, p. 19.

88 Ms Mia Garlick, Regional Director for Policy, Australia, New Zealand and the Pacific Islands, Meta, *Committee Hansard*, 20 January 2022, p. 20.

89 Giggle, *Submission 43*, pp 1-2..

90 Senator Claire Chandler, *Submission 69*, p. 1.

> The other thing I mentioned earlier was regarding being a woman. I've been a woman in a male-dominated field for many, many years. When I saw other people, potentially, talk out about this you would see so much commentary around it, that they were 'playing the victim' or 'playing the gender card'. It made me very reluctant to ever speak about this experience, because so much of the content and the things that were said to me had an angle that involved my gender. But the second you say that you get accused of playing that gender card, and that's something I never wanted to be accused of, I never wanted to do.[91]

Public versus private – abuse is abuse

3.74 A long trail of trauma, emotional suffering, reputational damage and sense of shame can be left regardless of whether abuse is directed towards someone with a public profile or not. Chad Wingard noted he felt how online trolls viewed him as a public personality and AFL player, rather than a person who played professional AFL as a job: 'A lot of people say that that comes with being an AFL player. But being bullied or discriminated against is not in the job description'.[92]

3.75 The National Mental Health Commission (NMHC) outlined that the impact of abuse is not lessened because an individual is well-known or because they have a prominent job or title:

> I go back to the point that the research is showing, which is that, when you are looking at issues of harm, it is highly specific to the individual and their usage. It's very highly specific. Of course, we all have vulnerabilities. We all have strengths. We all have different antennae. So I think it is very complex in that way. That would be my point. As I say, that's why, if I were looking at a cultural shift here, I would move very strongly to a 'do no harm' space and then look at what we need to do to shift towards there.[93]

3.76 The allowance of this behaviour, coupled with the reasoning from Big Tech for the two sets of standards is of concern to this Committee. Many witnesses cited free-speech, context or legitimate dissent or disagreement as the reason to allow abusive content to stay online.

---

91 Ms Erin Molan, *Committee Hansard*, 18 January 2022, pp 3-4.

92 Mr Chad Wingard, AFL, *Committee Hansard*, 1 February 2022, p. 24.

93 Ms Christine Morgan, Chief Executive Officer and Prime Minister's National Suicide Prevention Adviser, National Mental Health Commission (NMHC), *Committee Hansard*, 21 January 2022, p. 7.

> Social media corporations regularly fail to take action over direct threats of violence, wishes of harm against women, and threats of sexual assault against women. In some cases, this content is left online by social media companies despite being in breach of their own policies, while substantial effort is placed into moderating or banning other content which contains no threats or abuse.[94]

3.77 Ensuring there is a consistent standard applied by social media platforms when removing harmful or abusive content online will assist in reducing the proliferation of online harm and help to drive cultural change and improve the standard of public discourse online. This view was shared by various witnesses:

> It feels that the disconnect has been that social media companies have written their own rules, where other publishers and businesses that are disseminating information and creating content have a different set of rules and standards in the community and the wider sector that they have to obviously work within.[95]

> The impact of sexist, misogynist, 'gendered hate speech' (GHS) attacks or abuse of women, both in social media and online, have a significant impact. D'Souza et al state that: "The direct effects of GHS on the individual targets are neither trivial nor inconsequential. GHS has lasting impacts on women in terms of both their mental health and ability to participate in society free from fear.[96]

Online hate

3.78 Hate, including discrimination and hate speech, on online platforms was put as a significant issue facing users. As outlined in Chapter 2, users from particular backgrounds are more likely to experience discrimination and hate speech, including those from culturally and linguistically diverse (CALD) backgrounds, women, people with disability, migrant and refugee groups, and Aboriginal and Torres Strait Islander peoples.

3.79 Twitter stated in its submission that it had adopted a much broader definition of hate speech than currently exists in the *Racial Discrimination Act 1975* (Cth) (RDA). It stated that the RDA currently limits hate speech in relation to racial discrimination only, while Twitter had opted to include

---

94 Senator Claire Chandler, *Submission 69*, p. 2.

95 Mr Matt Berriman, Chair, Mental Health Australia, *Committee Hansard*, 21 January 2022, p. 24.

96 Ms Nicolle Flint MP, *Submission 70*, p. 3.

other categories of hate speech, such as in relation to sexual orientation, disabilities, and gender.[97]

3.80 Twitter further explained that it had conducted public consultation in relation to some of its policy updates, which informed some of these changes:

> That recommendation … came about during our dehumanisation updates to our hateful conduct policy. We had a multistaged approach for updating our hateful conduct policy that stepped through a number of marginalised communities, vulnerable communities and areas where we were seeing the contours of the online conversation had changed and were starting to not meet up with community expectations, where there was the ability for people to control their own experience or to report and have taken down content from Twitter. During these dehumanisation consultations we worked with a number of organisations here in Australia … We received a lot of feedback that, while there was hateful conduct being directed to certain individuals, it was the group conversations and the chronic abuse that vulnerable communities were suffering that was most harmful to them. This was directly inputted into the update that we had around racism and national origin and ethnicity updates to our dehumanisation policy.[98]

3.81 Twitter argued its Terms of Service and Rules are clear in relation to how it views discrimination:

> With regards to our hateful conduct policy, we are committed to combating abuse motivated by hatred, prejudice or intolerance, particularly abuse that seeks to silence the voices of those who have been historically marginalised. The policy makes clear that no one on Twitter may promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. We also do not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories.[99]

3.82 Meta similarly argued that its definition of hate speech was significantly broader than the definition provided by Australian legislation. Meta stated that it defines hate speech as 'a direct attack against people on the basis of what we call protected characteristics', which include race, ethnicity, place of

---

97 Ms Kara Hinesley, Twitter, *Committee Hansard*, 21 January 2022, p. 12

98 Ms Kara Hinesley, Twitter, *Committee Hansard*, 21 January 2022, p. 12.

99 Twitter, *Submission 50*, p. 8.

origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease.[100] 'Attacks' were defined by Meta as:

> violent or dehumanising speech, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, cursing, and calls for exclusion or segregation.[101]

3.83 Meta's latest Community Standards Enforcement Report for the period between July to September 2021 indicated that the company took action against 22.3 million forms of content including hate speech, 96.5 per cent of which was detected before it was viewed by people.[102] Meta also indicated that the prevalence of hate speech on its platforms was 10 to 11 containing hate speech out of 10,000 views of items of content, a rate of roughly 0.7 to 0.8 per cent.[103] Meta noted that hate speech was recognised as being exceptionally difficult to track via AI as it is 'dependent on nuance, history, language, religion and changing cultural norms'.[104]

3.84 Nonetheless, witnesses raised concerns regarding Meta's commitment to identify and remove hate speech. Ms Frances Haugen, former Facebook employee, stated that Meta could identify and remove significantly more content if it chose to invest resources appropriately. She said:

> A thing that might not be perfectly obvious is why Facebook takes down so little hate speech, and doesn't even take down hate speech in that many languages. There are 5,000 languages in the world, and a lot of the most fragile places are linguistically diverse. There are also spelling differences. American English is not the only version of English in the world, as you know. AI is not very smart. If you invest enough effort and you use the right techniques you can get very precise classification systems. But you're always forced to trade off between what fraction of the things do you want to catch and how often do you want your judgement to be wrong? Facebook has tried to be conscientious and avoid taking out content that is not violating. I think they genuinely value freedom of speech a great deal. But, when you look at those trade-offs, the trust they experience today at their current level of investment, they could invest more and take down more and still be more accurate and make fewer mistakes. But, if Facebook has to choose based on the systems that it has today, how much of the hate speech that exists do they want to take down and are

100 Meta, *Submission 49*, p. 43.

101 Meta, *Submission 49*, p. 43.

102 Meta, *Submission 49*, p. 44.

103 Meta, *Submission 49*, p. 45.

104 Meta, *Submission 29*, p. 44.

> they are willing to be wrong one in 10 times, one in 100 times, one in five times? They can take down more if they're willing to make more of those mistakes.[105]

3.85 Ms Haugen also noted that a considered understanding of what constitutes hate speech must accommodate significant amounts of subtly and nuance, which is extremely difficult for AI to capture. Using an example of a statement expressing that 'white paint colours are the worst' at a hardware store, she explained that a person would understand that the context and recognise that it was not hate speech, which a computer would not identify.[106] Ms Haugen argued that an example such as this demonstrated that removing harmful content was not necessarily practical, whereas reducing the systematic amplification of harmful content was more achievable.[107]

Algorithms on social media and other digital services

3.86 Algorithms are regularly cited as a digital feature used by many platforms which can be harmful. Meta described algorithms simply as 'just a set of rules that help computers and other machine-learning models make decisions', which are used in its systems to 'rank and distribute content' such as in its News Feeds on Facebook.[108]

3.87 Home Affairs defined algorithms in the context of the digital industry as being 'used to selectively predict the information that a user is more likely to engage with based on information about the user, such as location, past click-behaviour and search history'.[109] Home Affairs explained that algorithms are primarily used on social media to 'target users with content that appeals to their interests (filter bubbles)'.[110]

3.88 Google Australia explains why it uses algorithms and how it works on its platforms:

> With the vast amount of information available, finding what our users need would be nearly impossible without some help sorting through it. Google's

---

105 Ms Frances Haugen, *Committee Hansard*, 3 February 2022, p. 2.

106 Ms Frances Haugen, *Committee Hansard*, 3 February 2022, pp 2-3.

107 Ms Frances Haugen, *Committee Hansard*, 3 February 2022, p. 3.

108 Meta, *Submission 49*, p. 57.

109 Department of Home Affairs (Home Affairs), *Submission 40*, p. 4.

110 Home Affairs, *Submission 40*, p. 4.

> ranking systems are designed to do just that: sort through hundreds of billions of web pages and other content in our Search index to present the most relevant, useful results in a fraction of a second.
>
> To give users the most useful information, Search algorithms look at many factors and signals, including the words of their queries, relevance and usability of pages, expertise of sources, the user's context, such as their location and language, and settings. The weight applied to each factor varies depending on the nature of their queries. For example, the freshness of the content plays a bigger role in answering queries about current news topics than it does about dictionary definitions.[111]

3.89 As an example, Google Australia explained that YouTube's algorithms take into consideration signals in relation to a user's account (e.g. their watch history, previous searches, and their location), and will be overridden by YouTube's signals to reduce recommendations of harmful material where necessary.[112] Similarly, Meta stated that it uses algorithms in a variety of ways in its platforms to filter and rank content, some of which are designed to identify and remove harmful content.[113]

3.90 Concerns have been raised in relation to the ways in which algorithms can 'up score' negative or dangerous content. Home Affairs outlined its concerns that the selective prediction technology used by algorithms effectively encourage users to be isolated within particular viewpoints or ideologies, which can 'fuel racism, violence, and extreme political and/or narratives'.[114]

3.91 Home Affairs explained that algorithms utilised 'persuasive technology', which provides psychological rewards for engaging with the platform and ultimately encourages users to become reliant on the technology. It stated that persuasive design techniques are:

> design approaches conceived by human developers to maximise a user's engagement, usually tapping into social rewards and psychological behaviour, such as social reciprocity on social media platforms; pull-to-refresh content; or "streak rewards" in games or social messaging platforms. These designs reinforce and reward behaviours, such as people "liking" or "sharing" content. By expanding this and "pushing" content at users via notifications

---

111 Google Australia, *Submission 30*, p. 15.

112 Google Australia, *Submission 30*, p. 15.

113 Meta, *Submission 49*, p. 57.

114 Home Affairs, *Submission 40*, p. 4.

> and reminders, people are being conditioned to constantly engage with and even rely on these platforms and services, becoming digitally dependant.[115]

3.92 Reset Australia highlighted the three forms of harms that can flow from unmoderated algorithms in content prioritisation:

- societal harm (the over-promotion of divisive content);
- community harm (the tendency towards racism, sexism and other forms of discrimination); and
- Individual harm (including the normalisation of potentially harmful content).[116]

3.93 The Centre for Digital Wellbeing made a similar point, noting that algorithm-driven social media is creating harm at both the individual and societal levels:

> The impact of such platforms exists at both the individual level and the aggregate societal level. It is important to see the effect of social media on individuals as a continuum. Much of the focus of regulation has been on behavioural events, such as bullying, harassment or digitally enabled abuse. But, for a generation, the risks of usage include addiction, depression and anxiety. These impacts appear to be gendered, with young women most at risk.
>
> At a societal level, we are only beginning to understand the full influence of mass algorithmic engagement, but it is clear that social media is increasing polarisation and division. Such platforms are also the perfect conduit for the spread of misinformation and disinformation. The potential societal impacts are perhaps the most profound, yet to date much of the regulatory approach has focused on individual harms.[117]

3.94 Individual harm can also flourish in other ways due to algorithms. An example of harm online was raised by Eating Disorders Families Australia, which highlighted how pro-eating disorder content was easily found online on social media services, and often suggested by the platforms' algorithms which make recommendations based on a user's interests. They pointed to TikTok's algorithm as particularly harmful, arguing that the platform

---

115 Home Affairs, *Submission 40*, p. 4.

116 Reset Australia, *Submission 12*, pp 20-21.

117 Ms Carla Wilshire, Chair, Centre for Digital Wellbeing (CDW), *Committee Hansard*, 21 January 2022, p. 28.

tended to target young people's accounts in recommending 'diet' content, and would offer increasingly extreme content over time.[118]

3.95 Eating Disorders Families Australia explained the impact that algorithms promoting harmful content could have on vulnerable users:

> For example, if a user watches a "pro ana" (i.e., pro Anorexia) or "pro mia" (i.e. pro Bulimia) video, then they are likely to be supplied with more weight loss and "thinspo" (i.e. content to encourage them to lose weight) content, again resulting in validating and triggering behaviour which is known to intensify the deleterious impact of eating disorders. The impact of these social media sites is exacerbated by the fact they are visual and comparative in nature as well as encouraging users to be competitive in their postings, all of which are inherently problematic for young people battling with eating disorders.[119]

3.96 Witnesses also pointed to evidence suggesting that social media platforms are often aware of the harm algorithms cause but are unwilling to address it. Home Affairs stated that digital companies may be aware of the harm that algorithms cause in terms of promoting harmful or extreme content, but that this is overlooked in the interests of increasing viewership and, resultingly, revenue.[120] Platforms such as YouTube have also been found to have significant issues in relation to the way they use algorithms. The Australia Institute observed that, while there are documented harms in relation to the radicalisation of individuals via YouTube, the company has been unwilling to explain exactly how the algorithms work or provide access to enable broader understanding, and thus this remains an issue.[121]

3.97 Ms Frances Haugen stated that the social media business, and in particular Meta, is heavily reliant on large numbers of users generating and consuming content in order to maximise its profitability. She stated:

> Right now, Facebook is dependent on very, very large groups, like 500,000, groups of millions of persons, to fill people's newsfeeds with enough content. The amount of content people were producing on Facebook when it was just about their friends and families was enough that people could get online, spend 30 minutes, an hour, catch up with their friends and then go and do something else with their lives. If Facebook wants to make more and more

---

118 Eating Disorders Families Australia, *Submission 37*, p. 3.

119 Eating Disorders Families Australia, *Submission 37*, p. 4.

120 Home Affairs, *Submission 40*, p. 4.

121 The Australia Institute, *Submission 6*, p. 6.

> money every year, they have to keep you on their site longer and longer. The business model becomes the problem then, right? Once you start being dependent on these hyperamplification notes—when you have a group that has half a million people, a million people, it's not: say something offensive and there are 20 people in the room, with at most 20 people seeing that content. If I say something offensive in a room that has a few million people, the algorithm has a bias that the more extreme the content is, the more people it will reach. Suddenly, the thoughtful response to my thing doesn't get shown to two million people. That's not extreme enough. But my extreme thing goes out to two million people.[122]

3.98 The Australia Institute pointed to the example contained in documents, leaked by Ms Haugen, which described how Facebook's algorithms 'promoted posts which provoked angry reactions, as they generated more engagement than those which generated positive or neutral reactions'.[123] The papers suggested that employees tried to resolve this issue but was prevented from doing so by senior management, including Meta CEO Mark Zuckerberg, citing concerns that 'any intervention would lead to less engagement'.[124] Further, The Australia Institute noted that platforms such as Facebook were said to be overly dependent on AI technology rather than human detection of harmful content.[125]

3.99 Dr Salter explained that algorithms effectively suggest material to paedophiles on social media platforms (particularly YouTube) because it detects groups of users who are consistently looking at sexually explicit material of children and makes recommendations for similar videos. He put the view that this creates an 'alternate reality', where paedophiles are being promoted sexually explicit content of children which are not visible to others.[126]

3.100 Further, Dr Salter said because many social media platforms are reliant on users reporting harmful content, users that are accessing sexually explicit material of children are unlikely to report it, which creates a void space

---

122 Ms Frances Haugen, *Committee Hansard*, 3 February 2022, p. 3.

123 The Australia Institute, *Submission 6*, p. 5.

124 The Australia Institute, *Submission 6*, p. 5.

125 The Australia Institute, *Submission 6*, p. 5.

126 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 11.

where the content is not being monitored.[127] He provided an example of this situation in the context of YouTube, previously discussed in Chapter 2:

> … on YouTube the algorithm was detecting a group of users that were only looking at this potentially sexually explicit content of children, and the YouTube algorithm was then automatically generating a playlist of videos of children dancing, children doing gymnastics and children in swimwear. So, effectively, the algorithm has curated a paedophile playlist that is only visible to paedophiles … Far too often, what we're seeing occur on social media—and this has been true on Twitter; it has been true on other platforms—is that the social media algorithm is creating the parallel social media universe for child abusers, but the social media response system is reactive: it requires users to detect and report inappropriate behaviour. Well, if you're part of a community of child abusers on YouTube or Twitter or TikTok or wherever, you're not going to self-report the problematic content. So they're moving into these algorithmically-created stratums of inappropriate child content that the rest of us actually can't see. They're in a sort of a parallel universe.[128]

3.101 Algorithms were said to encourage users to access material that matches their worldview, effectively encouraging narrow-mindedness which can lead to intolerance. Professor Third suggested that algorithms can perpetuate discrimination due to their promotion of views and opinions that match the user's own rather than a more diverse range.[129] She explained that providing variety to users is essential for societal wellbeing:

> One of the big challenges I see is that it's very easy for children and young people to go online to be exposed to views that are of the same kind. It's really important not just for each individual to grow up and see themselves in digital media, or for them to learn about other cultures and so on, but for the health of our democracy that our children receive very diverse forms of information via a variety of channels and platforms. I know there's some very interesting work underway by some of the platforms to think about how we can use algorithms creatively to make sure we are serving up a diverse media diet. By increasing the diversity of not just children's but also adults' diets—you'll hear a regular theme here that some of the things we need to do for children also need to be

127 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 11.

128 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 11.

129 Professor Amanda Third, Professorial Research Fellow, Institute for Culture and Society, Western Sydney University; Co-Director, Young and Resilient Research Centre, Western Sydney University (Young and Resilient Research Centre), *Committee Hansard*, 21 December 2021, p. 30.

> done for adults—we are teaching them to understand people and giving them opportunities to be empathic about other people's experiences and so on.[130]

3.102 A similar point was also made by Ms Haugen, who used the example of a study into Instagram's algorithms in relation to eating disorders to demonstrate how the technology can lead young people towards harmful behaviours:

> You can go and follow moderate interests like healthy eating and healthy recipes on Instagram and just click on the content they provide you and you will get pushed towards eating disorder and self-harm content. Do it each day. Click on the first 10 things. Wait till the next day and do it again. You'll be shocked how fast the algorithm pushes you to more and more extreme ideas. If you were someone who was feeling depressed—you're a teenager, you're 16 years old, you're kind of struggling in school; maybe you're feeling kind of awkward and you start self-soothing by consuming Instagram. I think it's called doomscrolling, you're feeling stressed, so you just keep scrolling. If the content is what's making you depressed, that is dangerous, and saying that the solution to having an addictive product is having a tool that the addict has to pick to enable, that doesn't seem like a scalable solution.[131]

Addressing harm from algorithms

3.103 Some witnesses cautioned that algorithms are not inherently dangerous and can have positive functions if used appropriately. Home Affairs stated that many digital platforms currently already use algorithms to identify and stop child grooming conversations. It noted, however, that recent studies indicated that less than 40 per cent of companies in its survey utilised that form of technology to detect child exploitation. Further, Facebook's Friend Finder function was said to have been 'exploited by child sexual abuse live streaming facilitators to connect them with offenders'.[132]

3.104 In addressing concerns regarding algorithms, Meta has issued Content Distribution Guidelines which provide information on the forms of content that do not violate Community Standards but will not be distributed prominently due to problematic or low-quality content.[133] Further, Meta stated that it has created several features to enhance transparency for users,

---

130 Professor Amanda Third, Young and Resilient Research Centre, *Committee Hansard*, 21 December 2021, p. 30.

131 Ms Frances Haugen, *Committee Hansard*, 3 February 2022, p. 6.

132 Home Affairs, *Submission 40*, p. 5.

133 Meta, *Submission 49*, p. 3.

such as the 'Why Am I Seeing This?' function, and increased control over algorithms.[134]

3.105 More broadly, Meta observed that critics of social media algorithms argue that such systems are designed to encourage provocative or sensationalist content. In addition to arguing that its systems do not reward such content (and often work to reduce it), Meta further stated:

> The reality is, that it is not in Meta's interest — financially or reputationally — to continually turn up the temperature and push users towards ever more extreme content. The company's long-term growth will be best served if people continue to use its products for years to come. If our company prioritised keeping people online an extra 10 or 20 minutes, but in doing so made them less likely to return in the future, it would be self-defeating.
>
> Additionally, the vast majority of Facebook's revenue comes from advertising. Advertisers don't want their brands and products displayed next to extreme or hateful content.[135]

3.106 Twitter disagreed with the idea that algorithms are inherently problematic, commenting:

> Algorithm amplification is not problematic by default – all algorithms amplify. Algorithmic amplification is problematic if there is preferential treatment as a function of how the algorithm is constructed versus the interactions people have with it.[136]

3.107 Notwithstanding this point, Twitter stated further that it had made changes to its algorithms over time, enabling users to change the default setting so that users only see a reverse-chronology of Tweets rather than the algorithm-generated ranking. It also indicated that further developments to its algorithms were being made, with the goal of providing more choice to users in terms of how they wish to see content displayed.[137]

Algorithm regulation as a blunt tool

3.108 A number of witnesses were critical of the idea of issuing strict or blanket regulation on algorithms, arguing that they had significant benefits.

---

134 Meta, *Submission 49*, p. 3.

135 Meta, *Submission 49*, p. 67.

136 Twitter, *Submission 50*, p. 6.

137 Twitter, *Submission 50*, p. 5.

DIGI stated that algorithms were critical for the function of many online businesses and platforms, such as online mapping and word processing.[138] Automatic flagging of harmful content can be a function of algorithms, with recent reports highlighting that 95 per cent of videos containing harmful content were detected on YouTube.[139] Further, detection algorithms are used by social media platforms in relation to user behaviour, such as Twitter's detection of end-users abusing or harassing others, or Meta's algorithms which can find indicators of users at risk of self-harm or suicide.[140]

3.109 Digital Rights Watch (DRW) cautioned that the regulation of algorithms, including automated content moderation, could cause harm to those who have their content or accounts removed mistakenly. DRW argued that most digital platforms generally use automated technology to moderate content, which may result in certain forms of innocuous content being flagged, such as content produced by minority groups.[141]

3.110 DRW noted that automated content moderation is effective for some forms of media, such as images or videos, but is yet to effectively moderate audio or text content. Harmful content that is a mixture of image, audio, video and text, they argued, may not be flagged by automated processes.[142]

3.111 Dr Michael Salter also defended the use of algorithms in certain circumstances, arguing that algorithms have the capacity to be used for positive purposes. He stated:

> Algorithms and algorithmic detection have a critical role to play in child protection and that role could certainly be expanded. For example, there are automated detection systems for child sexual abuse material. Technology companies are not obliged to use them, although they are very effective. There are also algorithmic detection systems for inappropriate words, word combinations, emoji uses that can be used to detect child grooming—for instance, through message functions. Again, the technology companies at the moment are not obliged to use those sorts of detection systems. There is a lot that could be done at the back end and at the design end to make these products child safe. All that I would ask is that any service delivered to an

138 DIGI, *Submission 46*, p. 17.

139 DIGI, *Submission 46*, p. 17.

140 DIGI, *Submission 46*, p. 18.

141 Digital Rights Watch (DRW), *Submission 23*, pp 11-12.

142 DRW, *Submission 23*, p. 13.

> Australian child, whether it's online or offline, does not come with the predicted risk of that child being raped or sexually abused.[143]

3.112 This point was also raised by the CRF, who suggested that it was within the power of social media companies to incorporate this, as they have previously demonstrated this capacity:

> We saw how quickly social media platforms removed news media overnight. Do you remember? The question I ask is: Why can't they do the same in relation to online child exploitation? Why can't there be an absolute focus to remove that content from their sites. They have the AI technology. They have the ability. This is a continuing question that I ask of these various different providers.[144]

End-to-end encryption

3.113 End-to-end encryption (E2EE) is defined by the Office of the eSafety Commissioner (eSafety) as 'a method of secure communication that allows only the people communicating with each other to read the messages, images or files being exchanged'.[145] It is a widely used feature of apps such as WhatsApp, Signal and Skype, which utilise messaging and VoIP telephony services.[146] Some platforms such as Meta have indicated an intention to move towards full encryption of its services.[147]

3.114 E2EE is well-recognised for its capacity to provide secure and private communications, but has significant risks associated with it. eSafety's assessment of the risks includes issues such as:

- Avoiding scrutiny by law enforcement agencies for illegal activities, including online child sexual abuse, as detection technology generally

---

143 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 12.

144 Ms Sonya Ryan, CRF, *Committee Hansard*, 21 December 2021, p. 9.

145 Office of the eSafety Commissioner, *End-to-end encryption trends and challenges – position statement*, 11 May 2020, available at: https://www.esafety.gov.au/industry/tech-trends-and-challenges/end-end-encryption (accessed 6 February 2022).

146 Office of the eSafety Commissioner, *End-to-end encryption trends and challenges – position statement*, 11 May 2020, available at: https://www.esafety.gov.au/industry/tech-trends-and-challenges/end-end-encryption (accessed 6 February 2022).

147 Office of the eSafety Commissioner, *End-to-end encryption trends and challenges – position statement*, 11 May 2020, available at: https://www.esafety.gov.au/industry/tech-trends-and-challenges/end-end-encryption (accessed 6 February 2022).

does not work on E2EE systems, which may facilitate the production and distribution of CSAM; and

- Creating difficulty for law enforcement in detecting CSAM if social media companies begin to offer more E2EE services.[148]

3.115 Home Affairs expressed similar criticisms of the technology, arguing that E2EE is increasingly being utilised by criminals in order to avoid detection by law enforcement. Home Affairs argued that the 'increasing normalisation of these technologies on digital platforms, including social media, is bringing Dark Web functionality to the mainstream'.[149] It further outlined its concerns about the proliferation of E2EE functionality:

> While strong encryption plays an important role in protecting user privacy and data, the use of this technology in some settings, particularly on platforms used by children, brings with it important public safety risks. The application of end-to-end encryption across social media messaging services – such as expansion beyond the current opt-in services proposed by Meta (including on platforms such as Messenger and Instagram Direct) – will provide predators with the ability to evade detection as they connect with multiple vulnerable children anywhere in the world and develop exploitative grooming relationships. The nature of end-to-end encryption means that not even Meta, as the hosting company, would be able to retrieve or view these messages in order to detect child abuse, even under a lawfully issued warrant. The anonymity afforded by end-to-end encryption not only enables predators to groom victims on a social media platform, it also allows these criminal to safely connect and share tactics on how to perpetrate child sexual abuse, share explicit images, arrange live streaming of child sexual abuse through facilitators in vulnerable countries and avoid law enforcement.[150]

3.116 Further, Home Affairs stated that, when it had raised concerns, social media companies such as Meta expressed 'a degree of seeming indifference to public safety imperatives, including in relation to children'.[151]

3.117 Dr Michael Salter similarly stated that E2EE poses a significant and dangerous threat to efforts addressing the detection of CSAM:

---

148 Office of the eSafety Commissioner, *End-to-end encryption trends and challenges – position statement*, 11 May 2020, available at: https://www.esafety.gov.au/industry/tech-trends-and-challenges/end-end-encryption (accessed 6 February 2022).

149 Home Affairs, *Submission 40*, p. 5.

150 Home Affairs, *Submission 40*, pp 5-6.

151 Home Affairs, *Submission 40*, p. 6.

> At the moment the policy settings not just in Australia but internationally effectively disincentivise social media companies from becoming aware of illegal or harmful content, because once they are aware we start to see the expansion of their legal exposure and their potential liability. And so, as a result, it becomes in the interests of social media companies to make it difficult for them to know about illegal content. When we look at the move of, say, Facebook towards end-to-end encryption of the Messenger function, and in fact Twitter has also articulated an interest in encrypting its direct message service, then effectively what this does is it creates a black box in which the service cannot see any illegal content that's being exchanged and as a result they're not legally liable for that content. There is no incentive for them to proactively seek out and remove that material. That's not just true of social media, it's true of file-hosting services, it's true of a range of electronic service providers. The sheer amount of child sexual abuse material is increasing every year in the order of 50 per cent simply because this material is proliferating across services that have no legal obligation to proactively detect and remove that content.[152]

3.118 The CRF similarly took issue with the use of E2EE to enable offenders to avoid detection, stating that law enforcement would be significantly hindered if social media companies were to move towards E2EE services:

> End-to-end encryption is skewed towards providing greater privacy to adults at the expense of safety for children. In 2019, Interpol joined a list of law enforcement agencies in arguing that criminals hide behind the technology and that technology companies should be doing more to grant law enforcement agencies access across these channels. Law enforcement would have an incredibly reduced capacity to identify online child sexual exploitation offences without platforms proactively reporting instances to the National Center for Missing & Exploited Children. With no technological exception to end-to-end encryption, the dehumanising abuse of children, who will be left as collateral damage, will continue undetected. Their abusers and the people who trade the images and videos of abuse will be protected.[153]

3.119 Yourtown raised a similar point, noting that E2EE technology is primarily concerned with providing privacy, which does not align with the need to detect CSAM, and arguing this contrast needed to be addressed prior to any rollout by social media companies towards E2EE:

> There is an argument that that affords privacy, but that needs to be balanced with minimising significant harms to children who may be subject to abuse

---

152 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 12.

153 Ms Sonya Ryan, CRF, *Committee Hansard*, 21 December 2021, p. 3.

> and exploitation by end-to-end encryption processes. There needs to be always the overarching principle that the best interests of the child should be paramount in any design considerations. With some of those more secretive applications, there should be an opportunity for law enforcement, or indeed the providers, to be able to access information if it's in the best interests of the child and it's going to respond to significant safety concerns related to that child.[154]

3.120 In addition, Home Affairs noted that the technology industry had demonstrated that it was possible to detect CSAM material on E2EE platforms, such as features rolled out by Apple in August 2021 utilising 'hash' technology, similar to that used by Google (see above). It stated, however, that privacy concerns had been raised in relation to this technology, and that this may stop digital platforms implementing these features if there is backlash from the community.[155]

3.121 The polarisation of the debate regarding E2EE was raised by Dr Salter, who expressed concern that many organisations and groups had the view that 'all online data should be a black box that cannot be penetrated by law enforcement or by government'.[156] Home Affairs also highlighted this issue, arguing that while the debate regarding E2EE had become increasingly polarised, it was not a question of an 'all or nothing' approach. It argued that the digital industry was capable of developing technology and software which could scan for CSAM and other types of harmful content, and pointed to the example of Apple's new NeuralHash technology which detects known CSAM on iOS devices before it is uploaded to iCloud Photos.[157]

3.122 In February 2020, the United Kingdom-based National Society for the Prevention of Cruelty to Children (the NSPCC), in conjunction with over 300 child protection experts internationally including Dr Salter, wrote an open letter to Meta CEO Mark Zuckerberg, outlining its concerns with certain Facebook and Instagram features currently being proposed. In particular, the NSPCC cautions against the use of encrypted messaging and integration

---

154 Ms Kathryn Mandla, Head, Advocacy and Research, yourtown, *Committee Hansard*, 21 December 2021, p. 33.

155 Home Affairs, *Submission 40*, p. 6.

156 Dr Michael Salter, *Committee Hansard*, 18 January 2022, pp 12-13.

157 Home Affairs, *Submission 40*, p. 6.

into large open platforms, which it states would increase risks to children due to heightened exposure to abusers. [158]

3.123 The NSPCC called for Meta to adopt five key measures:

- Investment in safety measures showing that children's safety will be protected in the event that E2EE is introduced, including a capacity for Meta to scan for child abuse images and intervene where abuse is detected in its products or services;
- Indicate a commitment to instigate a voluntary duty of care in relation to the protection of children when designing encryption functions;
- A consultation process with child protection experts, governments and law enforcements;
- Sharing data with governments and child protection experts to ensure risks have been mitigation and how abuse behaviours have been impacted by the introduction of such technology; and
- Delaying a rollout of E2EE on Meta platforms until all mitigation strategies have been tested and can adequately address the concerns of child protection experts. [159]

3.124 Further, Home Affairs advised the Committee that when concerns had been raised with the platforms regarding the risks posed by E2EE, the companies had demonstrated that their priority was privacy rather than harm mitigation:

> The Department has ongoing concerns that digital platforms are prioritising privacy to the detriment of public safety. … For example, end-to-end encryption provides limited advantages over and above network level encryption. In the case of Facebook Messenger for example, end-to-end encryption will only apply to the content of messages, which has less commercial value to the company. The Department understands that personal data, such as metadata and site and cookie tracking, could still be exploited by Meta for commercial purposes, in line with their business model.[160]

3.125 Conversely, some witnesses were strongly in favour of E2EE being implemented more broadly. DRW argued that E2EE is critically important

---

158 National Society for the Prevention of Cruelty to Children, 'Open Letter to Mark Zuckerberg', available at: https://www.nspcc.org.uk/globalassets/documents/policy/letter-to-mark-zuckerberg-february-2020.pdf

159 National Society for the Prevention of Cruelty to Children, 'Open Letter to Mark Zuckerberg', available at: https://www.nspcc.org.uk/globalassets/documents/policy/letter-to-mark-zuckerberg-february-2020.pdf

160 Home Affairs, *Submission 40*, p. 6.

for online safety. While acknowledging the concerns regarding CSAM, DRW highlighted that encryption 'provides everyone with digital security, and protects everyone from arbitrary surveillance by malicious actors and cybercrime (e.g. identity theft)'.[161] Further, encryption was framed by DRW as a means of providing protection to vulnerable online actors, such as survivors of family violence who utilise encryption to protect information about escape and safe relocation.[162]

Privacy, anonymity and online harm

3.126 A topic of concern for all users of social media is the question of how users' data is treated and protected by digital companies. A number of topics were presented to the Committee which touched on issues of privacy, anonymity and how these issues link with online harm. This section examines these issues.

Anonymity and pseudonymity of online abusers

3.127 Anonymity and pseudonymity are used by many people when engaging in online activity in order to hide their true identity. Methods of hiding one's identity include:

- Providing false or no information to other individuals online about their personal characteristics such as name, age, or gender;
- Using an anonymised or pseudonymous account with minimal information provided to digital services in order to obtain access; and
- Mimicking or stealing another person's information (including their image).

3.128 Witnesses were sharply divided in their views of online anonymity, depending on the issue in which witnesses contextualised their responses.

Arguments for anonymity/pseudonymity

3.129 For many people, anonymity online is a necessary border between their online and offline lives. It allows people, particularly vulnerable people, the opportunity to engage with others online without worrying about the offline impacts of their online presence and in many cases encourages and enhances online participation.

---

161 DRW, *Submission 23*, p. 14.

162 DRW, *Submission 23*, p. 14.

3.130 The NMHC noted that they were not aware of any research conducted in relation to anonymity online and its impact on social connectivity.[163] Nonetheless, it observed that anonymity could be used positively or negatively, like social media and online platforms in general. The NMHC further argued that one of the main attractive points for clients of services such as Lifeline or Kids Helpline was its assurance of anonymity, which operates as a 'critical entry point' into mental health care services.[164]

3.131 WESNET emphasised the importance of online anonymity to women in dangerous situations, noting that 'Many people who do not use their real names on social media may have legitimate, non-nefarious reasons, such as people fleeing domestic violence'. WESNET highlighted an article by Dr Belinda Barnet, senior lecturer in media and communications at Swinburne University, who warned that any attempt to remove or reduce online anonymity could ultimately reduce safety: 'anonymity is important to your physical safety. Attacking anonymity on social media won't stop trolling, but it'll put sections of our community in danger'.[165]

3.132 DRW argued that allowing internet users to hide their identity is a vital aspect of digital interaction, which promotes freedom of speech and personal autonomy and empowerment:

> On a societal level, anonymity and pseudonymity online play an essential role in the functionality of the free and open internet, and enable political speech online which is integral to a robust democracy. On an individual level, the ability to be anonymous or use a pseudonym allows people to exercise control and autonomy over their online identity, to uphold their privacy. Anonymity is often an essential tool to protect individual safety and wellbeing. Any attempt to reduce the ability for people to be anonymous or pseudonymous online would undermine the above factors, and likely lead to increased long-term harm.[166]

3.133 DRW outlined a number of other reasons why people would use anonymity online, including:

---

163 Ms Christine Morgan, Chief Executive Officer and Prime Minister's National Suicide Prevention Adviser, National Mental Health Commission (NMHC), *Committee Hansard*, 21 January 2022, p. 8.

164 Ms Lyndall Soper, Deputy Chief Executive Officer, NMHC, 21 January 2022, *Committee Hansard*, p. 8.

165 WESNET, *Submission 25*, p. 3.

166 DRW, *Submission 23*, p. 5.

- Building communities online, especially in relation to marginalised groups such as the LGBTQIA+ community, people with disabilities, and ethnic minorities;
- Seeking information in relation to stigmatised health conditions;
- Victim/survivors seeking help in relation to domestic and family violence; and
- People in a public-facing role (such as medical staff, social/youth workers, lawyers, teachers, and so on) who wish to have an online life without being tracked down or contacted in relation to their work.[167]

3.134 Ms Carly Findlay AM noted that anonymity is necessary for many people, while the use of real names does not lessen the likelihood of users abusing or harassing others:

> Not all anonymous people are terrible. A lot of them have to use a pseudonym because they need to be protected legally, or there might be another reason for that. There's that. I definitely found that people who put their name to it are kinder. Also I've seen that people who put their name to things are really hateful.[168]

3.135 The Australian Information and Privacy Commissioner, Ms Angelene Falk, emphasised that anonymity is a key online principle:

> If you're a person who's experienced domestic violence and are wishing to access support online being able to do so without using your real name will be a way of keeping your identity and location private, in turn ensuring your safety. So it's an important privacy feature. It can also be a safety feature.
>
> In terms of a difference of views as to how anonymity ought to be played out in the online environment, we know, as I said, that anonymity and pseudonymity can actually ensure safety in certain contexts and that that privacy right can enable other rights and freedoms to be exercised—like freedom of speech, freedom of association and so on. One of the things I know the community is concerned about is the proliferation of abusive content online and the ability of people to engage in an online environment using a pseudonym. Again, we are talking about some very complex social policy issues and how to strike that right balance. From a privacy perspective, it's an important privacy right and it ought only to be displaced where there's a real

---

167 DRW, *Submission 23*, p. 6.

168 Ms Carly Findlay AM, *Committee Hansard*, 22 December 2021, p. 5.

> evidence base that it's necessary, reasonable and proportionate to achieve some other policy objective.[169]

3.136 The right to anonymity online is included in the Australian Privacy Principles (APP): 'Individuals must have the option of not identifying themselves, or of using a pseudonym, when dealing with an APP entity in relation to a particular matter'.[170]

3.137 Similarly, the Attorney-General's Department highlighted that there are many reason for anonymity online, and the Australian Government considers it important to protect that where it does not lead to harms:

> … there's a fine line between disincentivising defamatory comments online and not seeking to have a chilling effect on legitimate and appropriate comments online, including legitimate comments made anonymously. It's not the intention of government to stop anonymous use of the internet or anonymous comments being made on the internet. There are many, many legitimate reasons why a user may want to be anonymous when making comments, and it is the government's position that that is appropriate, provided that the content and substance of those comments, in this context, don't become defamatory and, in other contexts, don't become the other sorts of online harms that the government would draw the line against.[171]

Arguments against anonymity/pseudonymity

3.138 In contrast to these views, some submitters expressed concern that anonymity or pseudonymity encourages or amplifies harm. The majority of witnesses who were against the practice of online anonymity or pseudonymity considered the matter primarily from the perspective of preventing CSAM or child exploitation material in general. Ms Sonya Ryan, CEO and Founder of the CRF stated:

> Essentially, it provides a veil for criminals to hide behind. This provides an environment to be able to exploit children in various different ways. When it comes to our young people, I think that the government should be doing everything in its power to lift that veil, particularly for law enforcement

---

169 Ms Angelene Falk, Information Commissioner and Privacy Commissioner, Office of the Australian Information Commissioner, *Committee Hansard*, 28 January 2022, pp 19-20.

170 Office of the Australian Information Commissioner, *Australian Privacy Principles*, available at: https://www.oaic.gov.au/privacy/australian-privacy-principles/read-the-australian-privacy-principles (accessed 9 February 2022).

171 Mr Michael Johnson, Assistant Secretary, Defamation Taskforce, Attorney-General's Department, *Committee Hansard*, 28 January 2022, p. 39.

> agencies. Of course that would need to be regulated, but, when it comes to the safety of a young person, all measures should be made available and less anonymity given for users in the online space that are accessing children. We are seeing, especially with the proposed end-to-end encryption, more focus on privacy than there is often on the protection and safety of young people.[172]

3.139 Following on from this, Ms Ryan was also supportive of the implementation of age verification to monitor children's access to online services, as she believed that this woud also enable identity verification to assist with the 'unmasking' of anonymous perpetrators.[173]

3.140 More broadly, other witnesses argued that online anonymity encourages abuse or harassment. For instance, Ms Erin Molan noted that the power online trolls have comes from their anonymity:

> But the personal impact of this on people—and we've seen people take their lives. We've seen kids try to take their lives. We've seen so many lives ruined by this kind of behaviour. It's not weak and it's not for the vulnerable. It's not for the people who aren't resilient. Strong people get absolutely annihilated and torn to shreds by this behaviour, by anonymous trolls. You take away their anonymity and you take away their power and, all of a sudden, it's a level playing field again. And that's what it needs to be.[174]

3.141 While acknowledging that online anonymity plays important roles, particularly for whistleblowers, Dolly's Dream contrasted that with the problems it can lead to, and the need to respond to those in a nuanced way:

> But when it's used purely for the purpose of trolling somebody, bullying somebody, abusing somebody—all of the things that we know are the worst of social media—there are some of the issues on which we hope platforms, service providers and regulators are able to at least have a meaningful conversation and work with others.[175]

Privacy protection and data storage practices

3.142 Privacy considerations, including how users' data is stored and used by social media and other digital companies, were highlighted as an issue by witnesses. It was put to the Committee that digital platforms do not provide

---

172 Ms Sonya Ryan, CRF, *Committee Hansard*, 21 December 2021, p. 5.

173 Ms Sonya Ryan, CRF, *Committee Hansard*, 21 December 2021, p. 6.

174 Ms Erin Molan, *Committee Hansard*, 18 January 2022, p. 4.

175 Mr Stephen Bendle, General Manager, Dolly's Dream, *Committee Hansard*, 27 January 2022, p. 2.

sufficient transparency in advising how they treat their users' data and protect their privacy, particularly for vulnerable users such as children. [176]

3.143 Professor Amanda Third asserted that children are conscious of the digital industry's possession of their data, but are encouraged by adults' behaviour to agree to lengthy terms and conditions without fully understanding them. Professor Third stated:

> Children, like adults, have become attuned to the ways their data is collected, stored, used and shared on. For children here in Australia but also internationally, the big concern is that they really don't understand how their data is collected. They know it's happening, but they don't know what's happening to it and when and how that data can be used. They feel like they have to sign very complicated terms and conditions when they sign up to use social media and other platforms, and they don't always understand the deals they're making. This sends a very destructive message because, on the one hand, we teach children through a whole range of programs—for example, around healthy relationships—that consent is hugely important to the proper functioning of the world, but, on the other hand, when it comes to their technology practices we more or less tell them that it doesn't matter whether or not you understand the terms and conditions you're signing up to; check the box and off you go.[177]

3.144 Meta argued that it had developed strong privacy and data management practices, consulting with experts, government and the broader sector to create effective privacy tools.[178] It stated that its Privacy Review process enables Meta to build every new product or feature it creates with privacy considerations as a paramount consideration, and provides customers with 'choices and transparency'.[179] When a Privacy Review is conducted, Meta explained that a broad range of teams examine a product to determine the strength of its privacy protections:

> During this review, cross-functional teams evaluate privacy risks associated with a project, and determine if there are any changes that need to happen before launch to control for those risks. This review considers whether a project meets our privacy expectations which include: purpose limitation, data

---

176 Professor Amanda Third, Young and Resilient Research Centre, *Committee Hansard*, 21 December 2021, p. 27.

177 Professor Amanda Third, Young and Resilient Research Centre, *Committee Hansard*, 21 December 2021, p. 27.

178 Meta, *Submission 49*, p. 78.

179 Meta, *Submission 49*, p. 78.

> minimisation, data retention, external data misuse, transparency and control, data access and management, fairness and accountability.[180]

3.145 In relation to data privacy in particular, Meta provided the following summary of its approaches:

> We take a multi-faceted approach to data security, focussing on areas as diverse as penetration testing, spam prevention, disrupting operations run by adversaries, data protection, and taking legal steps to respond to cyber attacks. We've invested significantly to ensure our network infrastructure is strong, secure and capable of enforcing strong encryption for billions of users. We use a combination of expert teams and automated technology to detect potential abuses of our services.[181]

3.146 Further, Meta stated that it follows five principles in relation to data security matters: using encryption and security to protect user data; refusing to provide 'back door' government access; ensuring robust policies in relation to government requests for user data; refusing compliance where Meta feels a government request is 'deficient'; and providing transparency by way of providing notifications to users in relation to requests for the information prior to any disclosure.[182]

3.147 Google has developed a privacy protection program across its platforms, particularly in relation to children. For users under the age of 18 years, Google has implemented default privacy settings such as:

- Disabled functionality for ads personalisation;
- Upload settings set to the most private available, which restricts content to be seen only by the user and whoever they choose; and
- Educational material and guidance for parents and children.[183]

3.148 Snap stated that it has implemented Privacy by Design principles in the build of its app, demonstrated in its lack of shared newsfeeds and promoting communication between two people or small groups only.[184]

3.149 Twitter stated that it has a Privacy Centre which provides users with more information about its privacy practices, which provides information in

---

180 Meta, *Submission 49*, p. 78.

181 Meta, *Submission 49*, p. 83.

182 Meta, *Submission 49*, pp 83-84.

183 Google Australia, *Submission 30*, p. 7.

184 Snap Inc., *Submission 16*, p. 1.

relation to Twitter's Privacy Policy and guidance for users to improve privacy settings.[185]

Age verification technology

3.150 Age verification (or age assurance) tools have been increasingly utilised by technology and social media platforms to prevent children accessing inappropriate material. Meta emphasised that age verification 'is not as easy as it sounds', noting that it is a complex challenge to the entire digital industry to adequately understand a user's age.[186]

3.151 Google Australia explained that it uses a variety of means to determine the age of users before the platform restricts age-inappropriate content. It stated that '[t]hese measures can be supplemented with additional steps that ensure that children interacting with services are being treated appropriately while also respecting data minimisation requirements'.[187]

3.152 Meta took a slightly divergent approach, indicating a recognition that some users were too young to be on its service:

> Meta takes a multi-layered approach to understanding someone's age - we want to keep people who are too young off of Facebook and Instagram, and make sure that those who are old enough receive the appropriate experience for their age.[188]

3.153 As an initial step, Meta requires the provision of a date of birth when registering for a new account (known as an age screen). It will refuse access for those under the age of 13 years, and places restrictions on attempts to enter different birthdates into the age screen to circumvent the possibility of underage users attempting to 'game' the system.[189] Recognising that some users may lie about their age to gain access, Meta has invested in AI to understand a user's real age. Signals used by Meta to detect a person's true age include examining posts wishing someone a happy birthday and the age written in comments, and linked accounts with different ages associated with them.[190]

---

185 Twitter, *Submission 50*, p. 28.

186 Meta, *Submission 49*, p. 30.

187 Google Australia, *Submission 30*, p. 9.

188 Meta, *Submission 49*, p. 30.

189 Meta, *Submission 49*, p. 30.

190 Meta, *Submission 49*, p. 30.

3.154 If Meta identifies accounts that appear to be owned by a person under 13 years, and that person cannot provide evidence of their age, the account is deleted. Meta stated that this policy led to the removal of over 2.6 million accounts on Facebook and 850,000 accounts on Instagram between July and September 2021 due to minimum age requirements not being met. [191]

3.155 Additional controls for young people on their platforms include:

- Encouragement of private accounts for young people with existing public accounts on Instagram;
- Controls on advertisements targeting young people under 18 years, including restricting advertisements based on interests, activities on the platform, or activity on other apps or websites;[192]
- Default account provisions for young people, including privacy settings set at high levels;
- Warnings for sensitive content that are permitted on the platforms for 'public interest, newsworthiness or free expression value, that may be disturbing or sensitive for some users'[193] (e.g. violent or graphic content that provides evidence of human rights violations);
- Restrictions on adults sending private messages to young people, including a safety notice function being sent to users;
- Developing technology to make it difficult for adults to find or follow young people, specifically by identifying accounts which are demonstrating suspicious behaviour (e.g. being reported or blocked by a young person) and not allowing young people's accounts to display.[194]

Views of age verification

3.156 Many witnesses to the inquiry expressed concern that age verification technology would negatively impact on users' privacy. DRW argued that the impact on privacy was at odds with reducing harm online. It stated:

> Most forms of age verification require the provision of additional personal information in order to be effective. Incentivising companies and government agencies to collect, use and store additional personal information in order to

191 Meta, *Submission 49*, p. 34.

192 Meta, *Submission 49*, p. 32.

193 Meta, *Submission 49*, p. 32.

194 Meta, *Submission 49*, pp 31-34.

> conduct age verification creates additional privacy and security risk, which in-turn can exacerbate online harms.[195]

3.157 In outlining the ways in which Google conducts age assurance processes, Google Australia expressed its view that 'hard identifiers' or third-party verification methods should only be used for 'content and services that are particularly risky for children as they have a detrimental impact on all users' ability to access content and services'.[196] Google Australia also suggested that stricter forms of verification also may impact vulnerable groups who may be unable to access the required forms of identification such as credit cards.[197] It expressed the view that:

> No age verification mechanism is 100% accurate, and the more accurate the mechanisms the more intrusive it likely is. Ensuring that we implement age-appropriate safeguards, while at the same time ensuring that our services remain private and accessible remains a complex challenge. It's a problem that we are committed to solving, but no one company will be able to address this alone. Age assurance models should follow a risk-based assessment and be implemented in a proportionate way, balancing the need for accuracy with the risk of limiting rightful access to information and impact on users' privacy.[198]

3.158 The Daniel Morcombe Foundation noted that while many social media companies have put in place age restrictions for their users, these limits were easily overcome by parents or older siblings signing children up for accounts themselves.[199]

3.159 This topic was considered in an inquiry by the House Standing Committee on Social Policy and Legal Affairs, which recommended the development and implementation of a mandatory age verification regime for online pornographic material, to be undertaken by the eSafety Commissioner.[200] It further recommended that the National Consumer Protection Framework for Online Wagering introduce a requirement that 'customers are not able to

---

195 DRW, *Submission 23*, p. 7.

196 Google Australia, *Submission 30*, p. 9.

197 Google Australia, *Submission 30*, p. 9.

198 Google Australia, *Submission 30*, p. 10.

199 Ms Tracey McAsey, Manager, Daniel Morcombe Foundation, *Committee Hansard*, 21 December 2021, p. 14.

200 House Standing Committee on Social Policy and Legal Affairs, *Protecting the age of innocence: Report of the inquiry into age verification for online wagering and online pornography*, February 2020, pp 71-72 (Recommendation 3).

use an online wagering service prior to verification of their age as 18 years or over'.[201] The Australian Government was supportive of a number of the recommendations made, and work has commenced to address the concerns raised.[202]

Parental control features

3.160 Many online products such as social media platforms utilise parental control technology as part of their services to enable parents to monitor and control what content their children view and interact with.

3.161 Multiple witnesses suggested these controls were problematic in their approach to safety and were thus questionably achieving their goals of protecting children online. Some witnesses pointed out that these systems assume that a parent has the technological understanding to effectively use the controls.

3.162 Family Zone highlighted that the current approach tends to shift responsibility of children's safety online onto parents, and encourages blame to be placed on parents in the event that children are harmed:

> Too frequently the exposure of children to harm online is blamed on parents. There appears to be a popular but entirely fallacious view that "parents don't care" or "parents need to do more".
>
> In our experience this is categorically not true and anyone who has attempted to navigate the pitfalls, complexity and challenges of keeping kids safe online would agree.[203]

3.163 Professor Amanda Third suggested that the underlying assumption of parental control technology is that children have parents who can assist in using technology, which may not be the case. She stated that children who do not have the benefit of having parents who are able and willing to teach

---

201 House Standing Committee on Social Policy and Legal Affairs, *Protecting the age of innocence: Report of the inquiry into age verification for online wagering and online pornography*, February 2020, pp 88-89 (Recommendation 4).

202 Australian Government, *Australian Government response to the House of Representatives Standing Committee on Social Policy and Legal Affairs report: Protecting the age of innocence*, 1 June 2021, available at: https://www.aph.gov.au/DocumentStore.ashx?id=7a1aa6f4-b43b-4687-8e42-6686ce350beb (accessed 7 March 2022).

203 Family Zone, *Submission 15*, p. 3.

them about technological use and monitor their activities, may instead seek guidance in other places.[204]

3.164 Similarly, Reset Australia noted that:

> While parental control tools are important, we do not believe that placing responsibility onto parents to manage online safety is the most effective solution. Platforms should be developing systems that are safe for children in the first instance.[205]

3.165 Reset Australia drew the analogy with industrial hazard reduction, where the focus is on eliminating hazards rather than simply placing barriers between users and the hazard.[206]

3.166 This point was supported by Dr Michael Salter, who suggested that allowing platforms to place the responsibility on parents to control their children's online access and habits was problematic in two key ways.

3.167 Firstly, Dr Salter stated that not all children have protective parents. He explained that this could be due to a range of reasons, including that parents 'might be working three jobs, their parents might be experiencing substance abuse issues or mental health issues or … incapacitated for another reason'.[207] He also noted that some children may not be cared for by their parents, such as in being in residential care or out-of-home care.[208] Secondly, Dr Salter noted that a significant proportion of CSAM content is made within the home by parents and family figures, which indicates that these carers may not have their child's best interests at heart.[209]

3.168 The Isolated Children's Parents' Association of Australia further highlighted the particular challenges faced by children from remote parts of the country, many of whom are educated in boarding schools and therefore do not have parents in proximity to help with and monitor their technology usage.[210] Similarly to Dr Salter and Professor Third, they highlighted that many

---

204 Professor Amanda Third, Young and Resilient Research Centre, *Committee Hansard*, 21 December 2021, p. 27.

205 Reset Australia, *Submission 12*, p. 24.

206 Reset Australia, *Submission 12*, p. 24.

207 Dr Michael Salter, *Committee Hansard,* 18 January 2022, p. 13.

208 Dr Michael Salter, *Committee Hansard,* 18 January 2022, p. 13.

209 Dr Michael Salter, *Committee Hansard,* 18 January 2022, p. 14.

210 The Isolated Children's Parents' Association of Australia, *Submission 67*, p. 1.

children do not necessarily have their parents present to appropriately guide them in using online services.

3.169 Responding to these concerns, Meta stated that parental controls are available for some of its services. Messenger Kids, launched in 2020, enables parents to set privacy and security controls while enabling young people to utilise the product.[211] Meta intends to bring in parental controls for Instagram, including the capacity to monitor and set limits on their children's usage and enabling young people to notify their parents if they report someone.[212]

Limited reporting requirements

3.170 Multiple witnesses voiced concerns regarding the lack of transparency and mandatory reporting requirements from social media companies in relation to online harms that exist on their platforms.

3.171 Dr Michael Salter stated that, due to the business model that social media companies operate in, digital platforms are generally disincentivised to provide transparency in relation to online harm as this may scare potential investors or advertisers.[213] While citing Meta as a positive example of a company which openly reports about the detection and action taken against harms such as CSAM, other digital companies provide no oversight:

> In comparison [to Meta], we see almost no notifications from Apple year on year. But there is no question that significant amounts of CSAM are being created and shared via iPhones and also through various file storage and Cloud storage facilities.[214]

3.172 Ms Ryan stated that the CRF had sought data from social media services, but had been consistently refused on the basis of user privacy.[215] She argued that any platform hosting children should be transparent with data in order to better protect underage users, and that all online entities should provide data to the 'appropriate agencies' if it assists in protecting children online.[216]

---

211 Meta, *Submission 49*, p. 19.

212 Meta, *Submission 49*, p. 19.

213 Dr Michael Salter, *Committee Hansard,* 18 January 2022, pp 12-13.

214 Dr Michael Salter, *Committee Hansard,* 18 January 2022, pp 12-13.

215 Ms Sonya Ryan, CRF, *Committee Hansard,* 21 December 2021, p. 5.

216 Ms Sonya Ryan, CRF, *Committee Hansard,* 21 December 2021, p. 6.

3.173 In response, Google Australia stated that since 2010 it has published transparency reports to provide the public with information on government requests for data. It has expanded these reports by including YouTube data, information on how Google is addressing CSAM, and a Community Guidelines Enforcement Report which details the platform's efforts to reduce harmful content.[217]

3.174 In addition to providing transparency about its operations, Meta also stated that it welcomed regulation of harmful content.[218] Meta argued that it had 'led the industry in developing transparency reports about content enforcement', especially regarding the prevalence of users viewing content in violation of Meta's policies, which was divided by the estimated rate of content views in total across the particular platform (e.g. Facebook or Instagram).[219]

3.175 Meta also publishes Community Standards Enforcement Reports every quarter, which provide information in relation to the amount of content Meta removes or otherwise actions, the amount of content identified and removed prior to its being viewed.[220] Other transparency reports published by Meta include reports on widely viewed content, content restrictions, government and law enforcement cooperation, and other areas.[221] Meta's Oversight Board also publishes Transparency Reports which provide information about its deliberations and its decisions in relation to cases.[222]

3.176 Similarly, Twitter publishes biannual Transparency Reports detailing information such as enforcement of its Rules and foreign interference disruption. Twitter also has open access to its Transparency Centre which provides data in relation to issues such as information request, removal requests, copyright notices, and others.[223]

3.177 TikTok issues Transparency Reports on a quarterly basis in relation to harmful material and removal of content.[224] Snap publishes transparency

---

217 Google Australia, *Submission 30*, p. 17.

218 Meta, *Submission 49*, p. 71.

219 Meta, *Submission 49*, p. 9.

220 Meta, *Submission 49*, pp 72-27.

221 Meta, *Submission 49*, p. 71.

222 Meta, *Submission 49*, p. 77.

223 Twitter, *Submission 50*, p. 11.

224 TikTok Australia, *Submission 57*, p. 3.

reports on a biannual basis, and is currently the only major social media platform to break down statistics by country.[225]

Committee comment

3.178 In an ideal online world, technology would act as a neutral means of connecting people around the world, enabling safe and secure communication. Given the findings contained in this chapter, however, it is clear that some elements of digital platforms have the capacity to amplify harm and cause further distress to victims.

3.179 The Committee notes the work that has been conducted by the social media companies particularly in attempting to promote safety for users on their platforms. It also commends the continuing efforts of technology companies to provide safety for users, particularly vulnerable users such as children.

3.180 Nonetheless, the Committee is of the view that there is much more for industry to do to assure governments and the public that platforms are taking the matter seriously and that Australians can trust them to protect their safety online.

System design

3.181 The Committee agrees with the views of witnesses that social media platforms and digital products at large have generally not been designed with users' safety as a priority, particularly for vulnerable groups.

3.182 The Committee is particularly mindful of Dr Michael Salter's comments in relation to this matter, where he compared the online industry to the childcare sector in the 1970s and 1980s, which developed exponentially and rapidly due to increased demand for the sector, but did not adequately consider the risks to children and as a result experienced significant issues in relation to child sexual abuse.[226] The online industry is in a similar space of development, where safety principles are now being accepted as necessary to ensure the safety of what is now an essential service for many Australians but are yet to be substantially implemented by providers.

3.183 The Committee commends the work of the eSafety Commissioner in establishing the Safety by Design Principles, and looks forward to seeing the implementation of these principles as the *Online Safety Act* comes into effect.

---

225 Snap Inc., *Submission 16*, p. 3.

226 Dr Michael Salter, *Committee Hansard*, 18 January 2022, p. 12.

The Committee is pleased to note that the social media and digital platforms appear to have been cooperative and constructive during consultations for the design of the Safety by Design Principles, and hopes to see their commitment to online safety turning into effective action on their platforms.

3.184 Having said that, while the Committee is cognisant of the work being conducted by social media companies in taking action on harm in relation to the Safety by Design principles, it notes that a number of social media companies and digital platforms have been in the industry for in excess of ten years, and that the Safety by Design principles will need to be retrospectively implemented. The Committee is eager to see how such retrospective application will be carried out, and believes that this topic would be best examined by the upcoming review of the operation of the *Online Safety Act*.

Recommendation 9

3.185 The Committee recommends that future reviews of the operation of the *Online Safety Act 2021* take into consideration the implementation of the Safety by Design Principles on major digital platforms, including social media services and long-standing platforms which require retrospective application of the Safety by Design Principles.

Identifying and removing harmful content

3.186 The Committee is concerned by the evidence provided regarding the harms that are present on digital platforms, as outlined in Chapter 2. These forms of harm are pernicious to those who experience them, resulting in extremely real and serious suffering, which in many cases has impacts which affect people's entire lives and are long-lasting.

3.187 Social media and other digital companies appear to be addressing these harms through a mixture of proactive detection technology and reports from users once content is published on the platform. While the Committee notes that social media companies are improving at their detecting of harmful content before it is made public, it remains concerned that these detection models are inherently reliant on responding to harmful content after it has been made public or has occurred, and do not go sufficiently far enough in preventing harm.

3.188 Another key issue raised by witnesses was the inadequacy of the industry's policing or moderation of their own standards. Many examples were given of abusive comments that met the threshold to involve either the police or

the eSafety Commissioner, but the comments were not taken down by the platforms themselves.

3.189 Platforms need to properly monitor and uphold their own standards. There should be clear and direct consequences for breaches, including (but not limited to):

- Banning users from social media platforms and other digital spaces; and
- The use of pop-up warnings for content that an algorithm identifies as potentially breaching terms of service.

3.190 Further, the Committee objects to the trend of many social media platforms that places too much responsibility for identifying harmful content on user reports. Everyday users are not technically or psychologically equipped to be able to identify harm at large and withstand the trauma that can potentially result. Social media users should be able to trust that their interactions online will be safe and free from harmful content which could potentially cause extreme and long-term effects on viewers.

3.191 The detection of CSAM content is of particular interest to the Committee. The inconsistent approaches of social media platforms and digital platforms in detecting and removing CSAM content indicates that further work is required to ensure that the industry does not inadvertently protect and facilitate perpetrators. Accordingly, it would be appropriate for digital platforms to investigate improved methods of detection in relation to CSAM material, in consultation with both Australian and international child exploitation authorities.

3.192 The Committee notes the comments by a number of witnesses in relation to the dangers posed by a mass uptake of E2EE by social media and digital companies. It is not clear that the risks posed by the potential shielding of CSAM (in addition to other forms of harm) over E2EE services are outweighed by the benefits offered by services to provide privacy to users. While these services can be used by vulnerable groups, such as those experiencing family violence, the Committee also considers that one vulnerable group's rights should not negate another group's rights to protection from harm. Any potential widespread uptake of E2EE should be carefully considered and – if necessary – regulated to ensure that the appropriate balance between harm reduction and privacy protection is maintained.

3.193 The Committee's view is that the challenges raised by E2EE in relation to the implementation by social media platforms are real and of concern. Without

corresponding safeguards, E2EE can negatively impact the ability of law enforcement agencies to identify online predators.

3.194 Additionally, it became increasingly clear during the course of the inquiry that while social media companies may have strong policies and terms of use in relation to harmful content on their services, they experience significant challenges in upholding these standards. When provided examples of abusive comments that met the threshold for eSafety's adult online abuse framework and also for police investigation, a number of social media platforms did not remove these comments or consider them to have breached the relevant terms of use.

3.195 The community standards that platforms apply to their users do not necessarily match up with general community expectations. Social media companies have a role in shaping societal standards regarding what is and is not appropriate behaviour online or in-person. The Committee is concerned that community standards are being shaped and influenced by social media platforms, and without regulatory intervention that online abuse will continue to flourish.

3.196 Accordingly, the Committee believes that the eSafety Commissioner should be given additional powers to compel social media companies to provide evidence for the enforcement of their terms of service and rules. It further is of the view that the eSafety Commissioner should be given greater powers to disrupt volumetric attacks of individual users.

Recommendation 10

3.197 The Committee recommends that the Department of Infrastructure, Transport, Regional Development and Communications, in conjunction with the eSafety Commissioner and the Department of Home Affairs, examine the need for potential regulation of end-to-end encryption technology in the context of harm prevention.

Recommendation 11

3.198 The Committee recommends that the eSafety Commissioner, as part of the drafting of new industry codes and implementation of the Basic Online Safety Expectations:

- **Examine the extent to which social media services adequately enforce their terms of service and community standards policies, including the**

efficacy and adequacy of actions against users who breach terms of service or community standards policies;

- **Examine the potential of implementing a requirement for social media services to effectively enforce their terms of service and community standards policies (including clear penalties or repercussions for breaches) as part of legislative frameworks governing social media platforms, with penalties for non-compliance; and**

- **Examine whether volumetric attacks may be mitigated by requiring social media platforms to maintain policies that prevent this type of abuse and that require platforms to report to the eSafety Commissioner on their operation.**

Recommendation 12

3.199 The Committee recommends that the eSafety Commissioner examine the extent to which social media companies actively apply different standards to victims of abuse depending on whether the victim is a public figure or requires a social media presence in the course of their employment, and provides options for a regulatory solution that could include additions to the Basic Online Safety Expectations.

Algorithms

3.200 While algorithms play a key role in the basic function of multiple types of online services, it is clear that they have the potential to enormously accentuate online harm. This is heightened by the evidence from witnesses suggesting that many social media platforms are opaque in explaining how their algorithms work.

3.201 Algorithms require further investigation to determine the types and scale of harm they can cause, how they operate in different digital mediums, and how best to moderate and regulate them.

3.202 Notwithstanding this point, evidence provided suggested that many social media companies do not provide publicly available detail in relation to how their algorithms work and whether the platforms are doing anything to address potential harms caused through their algorithms. While some of this material may be considered as commercial-in-confidence, more transparency is required of social media companies to demonstrate that these concerns are being addressed.

3.203 A statutory requirement ensuring that digital platforms and social media companies provide details regarding how they are working to minimise harm caused by algorithms would be an appropriate form in which to provide this detail, which could be designed to ensure that it would not impact on the platforms' commercial-in-confidence details. The review of the *Online Safety Act* should consider whether such a requirement could be implemented and the appropriate form for it to take.

Recommendation 13

3.204 The Committee recommends that the eSafety Commissioner, in conjunction with the Department of Infrastructure, Transport, Regional Development and Communications and the Department of Home Affairs and other technical experts as necessary, conduct a review of the use of algorithms in digital platforms, examining:

- **How algorithms operate on a variety of digital platforms and services;**
- **The types of harm and scale of harm that can be caused as a result of algorithm use;**
- **The transparency levels of platforms' content algorithms;**
- **The form in which regulation should take (if any); and**
- **A roadmap for Australian Government entities to build skills, expertise and methods for the next generation of technological regulation in order to develop a blueprint for the regulation of Artificial Intelligence and algorithms in relation to user and online safety, including an assessment of current capacities and resources.**

Recommendation 14

3.205 The Committee recommends that the eSafety Commissioner require social media and other digital platforms to report on the use of algorithms, detailing evidence of harm reduction tools and techniques to address online harm caused by algorithms. This could be achieved through the mechanisms provided by the Basic Online Safety Expectations framework and Safety By Design assessment tools, with the report being provided to the Australian Government to assist with further public policy formulation.

Algorithmic transparency

3.206 The Committee was told repeatedly by the platforms during the course of this inquiry that their practices for dealing with online harms have improved.

3.207 However, these claims were difficult to assess in the absence of consistent, specific and auditable transparency frameworks. The Committee is of the view that this makes measurement on such improvements over time very difficult.

3.208 As Ms Frances Haugen noted:

> (The) inability to see into Facebook's actual systems and confirm how they work, as communicated, is like the Department of Transportation regulating cars by only watching them drive down the highway.[227]

3.209 Former CEO of Crowdtangle and a former employee of Meta Mr Brad Silverman, outlined proposals to empower the United States Federal Trade Commission to enact tiers of transparency on social media platforms. This proposed legislation would utilise three transparency mechanisms:

> The first is a mechanism for allowing in-depth research on very sensitive datasets that would be facilitated by an agency that sits under the National Science Foundation. The idea would be that academics and researchers would propose doing research on certain types of datasets. The NSF would function as a mediator…
>
> The second piece is what's called a safe harbour, which would allow public interest and news gathering related use cases for automated collection of data off platforms. That is colloquially oftentimes referred to as scraping, but it would certify the rights of certain entities to do automatic collection off public datasets…
>
> The third one is…essentially a handful of different other mechanisms [such as] a set of libraries designed to provide access to datasets that the public, writ large, could look at.[228]

3.210 Ms Frances Haugen made reference to the 'floor for transparency' and said that:

---

227 *Committee Hansard*, 3 February 2022, p. 6.

228 Mr Brandon Silverman, *Committee Hansard*, 28 January 2022, pp 43-44.

> Any system that we put in place needs to be dynamic. It needs to be a thing where the threats that we are aware of with Facebook will be different six months or a year from now. ... We need a system that is not about: 'Here are the 10 … things each day or each week or each year.' We need systems that are dynamic and respond to emergent concerns.[229]

3.211 Developing an effective transparency framework for the social media platforms is a complex challenge for policy makers and regulators, but an important one to address. In many senses it is the policy intervention that the success of many other interventions rests upon.

3.212 The Committee considers that requiring social media platforms to be transparent is a complex challenge for policy makers and regulators but it is an important one.

Recommendation 15

3.213 The Committee recommends that, subject to Recommendation 19, the proposed Digital Safety Review make recommendations to the Australian Government on potential proposals for mandating platform transparency.

Protection of children

3.214 The Committee acknowledges the work of the social media platforms and digital services to protect children while on their sites. Having said that, it is clear from evidence that children and their safety have not been considered as a primary focus when designing products, nor in managing users' experiences. Regardless of the intentions of social media platforms in attempting to keep underage users off their services, the reality is that children will inevitably access these services, and some to their detriment. Social media platforms have a fundamental moral duty to ensure that children are kept safe on their platform, regardless of age restrictions.

3.215 Children have the right to their privacy for their early years. This, however, is not a right being strictly protected by social media platforms at present. Strong protections should be considered as a paramount requirement for children, and privacy settings on social media accounts should be set as high as possible for those under 18 years of age

3.216 Further, the broader technological industry has a role to play in the battle to ensure children remain safe online. Additional requirements on technology

---

229 Ms Frances Haugen, *Committee Hansard*, 3 February 2022, p. 6.

companies, particularly those who design and produce technological devices such as smartphones, are required to ensure that parents are able to effectively monitor and control their children's use of social media services, and intervene before harmful situations arise.

3.217 The Basic Online Safety Expectation Determination includes that industry should make sure the default privacy and safety settings of services targeted at or used by children are robust and set to the most restrictive level. Notwithstanding this, the Committee considers there is an opportunity to build on this determination by making it a mandatory requirement for all digital services with a social networking component to set default privacy and safety settings at their highest form for all users under 18 years of age.

Recommendation 16

3.218 The Committee recommends the implementation of a mandatory requirement for all digital services with a social networking component to set default privacy and safety settings at their highest form for all users under 18 (eighteen) years of age.

Recommendation 17

3.219 The Committee recommends the implementation of a mandatory requirement for all technology manufacturers and providers to ensure all digital devices sold contain optional parental control functionalities.

4. Policing the Trolls

Legal Protections and Government Action on Online Safety

Introduction

4.1 Australia is at the forefront in legislative responses to online harm. With the introduction of the *Online Safety Act 2021* (OSA), in addition to a suite of other measures which have been introduced over the past decade, Australian law has provided a strong model for legislatures around the world in regulating online harms.

4.2 This chapter outlines the key elements of Australia's online safety legislative regime and the associated Government policy measures, before considering international responses to the challenge of regulating online environments.

4.3 Regardless of these achievements, the online environment does not stand still for governments to create regulation. Developments in technology and platforms occur at a dizzying pace, and the present regulatory environment may not fit a digital world within even a few short months. This chapter also considers potential areas where Australia's regulatory framework may require further change to appropriately address online safety concerns.

Legislative framework governing online safety

4.4 This section outlines the recent history of legislative powers in relation to online safety. It examines the activities and powers of the Office of the

eSafety Commissioner, including the programs it conducts (such as educational tools and reporting schemes). It also considers the portfolio division of responsibility of these powers.

Legislative powers pre-2021

4.5 Prior to 2015, Australia's primary legislation governing online safety was the *Broadcasting Services Amendment (Online Services) Act 1999*.

4.6 In 2015, the Australian Government enacted the world-first *Enhancing Online Safety Act 2015* (EOS Act), which established the role of the eSafety Commissioner (first named the Children's eSafety Commissioner) and the Office of the eSafety Commissioner (eSafety).

4.7 eSafety's powers and functions were further expanded by the 2017 amendment of the EOS Act, particularly in relation to the increase in the eSafety Commissioner's remit to cover all Australians.[1] A civil penalties scheme was established in 2018 that enabled eSafety to assist with the removal of intimate images or videos from online platforms, including in some cases to take action against the perpetrator, social media services, websites or hosting providers.[2]

4.8 eSafety has additional powers and functions in relation to certain categories and forms of content under other legislation, such as:

- the *Broadcasting Services Act 1992* (schedules 5 and 7);
- the *Telecommunications Act 1997* (section 581(2A));
- the *Enhancing Online Safety (Protecting Australians from Terrorist or Violent Criminal Material) Legislative Rule 2019* (section 5); and
- the *Criminal Code Act 1995* (sections 474.35 and 474.36).

Online Safety Act 2021

4.9 The EOS Act was superseded by the *Online Safety Act 2021* (OSA). The OSA was passed by the Parliament in June 2021 and came into force on 23 January 2022.

4.10 The intention of the OSA, according to Infrastructure, is that 'the rules and protections Australians enjoy offline should also apply online … and the Act provides a safety net for people when things go wrong online'.[3]

---

1 *Enhancing Online Safety for Children Amendment Act 2017* (Cth).

2 Infrastructure, *Submission 44*, p. 2.

3 Infrastructure, *Submission 44*, p. 3.

4.11 The OSA creates or updates reporting schemes to address cyber-bullying, adult cyber abuse, image-based abuse, illegal and restricted online content, and abhorrent violent content. [4] The OSA gives eSafety powers to require the removal of content reported under these schemes, with non-compliance attracting civil penalties directed to the poster of such material and provider of the service hosting the material. [5]

4.12 In addition, the OSA gives eSafety new information-gathering and investigative powers to assist in the administration of the schemes. It also permits eSafety to make Restricted Access System declarations, to control access to certain materials based on age.[6]

4.13 The OSA also has a mandatory independent review of its operation, set for three years after its commencement (23 January 2025). eSafety will also report on the operation of its schemes in its annual report.[7]

Reporting functions

4.14 The new legislation introduces or updates a number of reporting functions and schemes administered by eSafety, which are discussed below in detail.

Adult Cyber Abuse reporting scheme

4.15 The OSA establishes a new cyber abuse reporting scheme for adults, designed for Australians over the age of 18 years to report harmful conduct across a range of platforms. [8] The scheme requires that online service providers and/or individual users to remove cyber abuse targeting Australian adults 'with the intention of causing serious harm'. [9] The legislation stipulates that the threshold for what constitutes 'cyber abuse' is high and refers to the most serious forms of abuse, requiring two elements:

1 The abuse must be intended to cause 'serious harm', which means serious physical harm or serious harm to a person's mental health - like threats intended to cause serious psychological harm or serious distress that goes beyond ordinary fear

---

4 Infrastructure, *Submission 44*, pp. 3-6.

5 Infrastructure, *Submission 44*, pp. 3-6.

6 Infrastructure, *Submission 44*, pp. 3-6.

7 eSafety Commissioner, *Submission 53*, p. 6.

8 eSafety Commissioner, *Submission 53*, p. 6.

9 eSafety Commissioner, *Submission 53*, p. 21.

2 The abuse must also be menacing, harassing or offensive in all the circumstances.[10]

4.16 eSafety provided the following examples of abuse sufficient in meeting the adult scheme's threshold:

> publishing private or identifying information about an individual with malicious intent to cause serious harm; encouraging violence against a specific Australian adult based on their religion, race or sexuality; and threats of violence that make a person afraid they will suffer physical harm.[11]

4.17 It was noted that the threshold for adult cyber abuse is significantly higher than the threshold applied to the children's cyberbullying scheme. eSafety explained that this is due to the 'expectation that adults are generally more resilient than children', and to ensure the thresholds are consistent with those contained in the Criminal Code.[12] It did, however, note that it would monitor whether the legislative definition should be changed given the concerns about the high threshold.[13]

4.18 eSafety estimates that, since the commencement of the OSA and with investigations incomplete, less than 10 per cent of complaints meet the required threshold.[14] Further, as at 15 February 2022, eSafety had received over 200 complaints in relation to serious adult cyber abuse which met the requisite threshold.[15] Nonetheless, eSafety stated that it would deliver advice and guidance to all people seeking assistance with cyber abuse regardless of whether complaints met the threshold or not.[16]

4.19 Further, eSafety advised that certain topics, such as hate speech directed at groups as opposed to individuals, are beyond the scope of the new adult reporting scheme. It did confirm, however, that in assessing adult cyber abuse reports that eSafety would consider factors such as whether abuse was

---

10 eSafety Commissioner, *Submission 53*, p. 22.

11 eSafety Commissioner, *Submission 53*, p. 22.

12 eSafety Commissioner, *Submission 53*, p. 22.

13 eSafety Commissioner, *Submission 53*, p. 13.

14 eSafety Commissioner, *Submission 53.1*, p. 8.

15 Ms Julie Inman Grant, eSafety Commissioner, Office of the eSafety Commissioner, *Committee Hansard*, Senate Additional Estimates 2021-22, Senate Environment and Communications Legislation Committee, 15 February 2022, p. 23.

16 eSafety Commissioner, *Submission 53*, p. 22.

targeted at the individual in question on the basis of race or cultural background.[17]

Children's cyberbullying scheme

4.20 The cyberbullying scheme focused on children has also been updated in the OSA, reflecting the range of services outside of social media services where cyber bullying can occur. [18] eSafety explained that:

> whereas the previous scheme was limited to 14 specific social media services across two 'tiers' – one voluntary and one mandatory – the enhanced scheme will apply to all social media services, as well as a range of other services where cyberbullying can occur, such as messaging and gaming services.[19]

Image-based abuse reporting scheme

4.21 The image-based abuse scheme was also updated to enable eSafety to 'rapidly address the non-consensual sharing of intimate images'.[20] The pre-2021 scheme, first commenced in September 2018, enabled eSafety to investigate and act in response to complaints regarding the sharing or threatened sharing of intimate images without consent. [21]

4.22 The OSA provides minor but significant changes, such as:

- The introduction of a reduced time period (now set as 24 hours for all schemes) for online service providers to respond to a removal notice from the eSafety Commissioner, and
- New powers for eSafety to utilise a discretion to publicly name any service providers who consistently fail to manage online harms such as image-based abuse. [22]

Basic Online Safety Expectations

4.23 A key feature of the OSA is its creation of Basic Online Safety Expectations (BOSE). Under the BOSE, more responsibility will be placed on social media

---

17 eSafety Commissioner, *Submission 53*, p. 23.

18 eSafety Commissioner, *Submission 53*, p. 6.

19 eSafety Commissioner, *Submission 53*, pp 15-16.

20 eSafety Commissioner, *Submission 53*, p. 19.

21 eSafety Commissioner, *Submission 53*, p. 20.

22 eSafety Commissioner, *Submission 53*, p. 21.

and internet companies to provide a safe environment for their users.[23] The BOSE consists of:

- Core expectations, which are already laid out in the OSA itself;
- Additional expectations, which may be specified by the Minister via legislative instrument; and
- Reporting requirements, which will be imposed on social media and internet companies.[24]

4.24 Infrastructure stated that the BOSE was intended to set 'minimum safety expectations of online service providers, establishing a benchmark for online service providers to take proactive steps to protect the community from abusive conduct and harmful content online'.[25] Further, the BOSE establishes that:

> Providers of these services are expected to take steps to meet the Expectations included in the Determination and protect Australians from unlawful and harmful material and activity that falls within the remit of the enabling legislation the Online Safety Act 2021 (the Act), or impedes the online safety of Australians.[26]

4.25 As authorised under section 45 of the OSA, the BOSE was established on 20 January 2022 as a legislative instrument. The *Online Safety (Basic Online Safety Expectations) Determination 2022* (the BOSE Determination) sets out six key areas of expectations for social media and online platforms, which are divided into core expectations and additional expectations.
Table 4.1 provides a breakdown of the BOSE Determination's expectations.

4.26 The eSafety Commissioner, Ms Julie Inman Grant, stated that the OSA's establishment of the BOSE would assist in shifting the burden of responsibility of online safety towards social media platforms:

> The act is raising the bar on what government expects of the tech industry by introducing a basic set of online safety expectations, or the BOSE. Mandatory industry codes will also require the online industry to detect and remove illegal material, while preventing access to harmful content. Unlike [eSafety's]

---

23 Infrastructure, *Submission 44*, p. 3.

24 Infrastructure, *Submission 44*, p. 3.

25 Infrastructure, *Submission 44*, p. 3.

26 *Online Safety (Basic Online Safety Expectations) Determination 2022 – Explanatory Statement*, 23 January 2022, available at: <https://www.legislation.gov.au/Details/F2022L00062/Explanatory%20Statement/Text> (accessed 10 February 2022).

reporting schemes, the BOSE are not limited to specific forms of online harm and may enable us to shine a light on systemic failings and will compel transparency that is currently lacking, tackling issues such as online hate, self-harm content and the extent to which algorithms contribute to harm.[27]

Table 4.1 Basic Online Safety Expectations - breakdown of expectations by category

Division topic	Expectation/s or core expectation/s on providers	Additional expectation/s on providers
Safe use	Reasonable steps required to ensure safe use Consultation with the eSafety Commissioner and reference to eSafety's guidance in determining reasonable steps	Reasonable steps required regarding encrypted services Reasonable steps required regarding anonymous accounts Consultation and cooperation with other providers in promoting online safety
Treatment of certain content and activity	Reasonable steps required to minimise provision of certain content[28] Reasonable steps required to prevent access of class 2 material to children[29]	-
Reports and complaints	Provision of mechanisms to report and make complaints regarding certain content	Provision of terms of use, certain policies, etc.

27 Ms Julie Inman Grant, eSafety Commissioner, Office of the eSafety Commissioner, *Committee Hansard*, 3 February 2022, p. 12.

28 This includes: cyber-bullying material directed at an Australian child or adult, an non-consensual intimate image of a person, class 1 material, material promoting, inciting, instructing in or depicted abhorrent violent content.

29 Class 1 and Class 2 content are defined under the Online Content Scheme in the Act. Class 1 primarily refers to child exploitation material and pro-terrorism content, or offend the standards of morality, decency and propriety generally accepted by reasonable adults. Class 2 material is considered inappropriate for general public access or for children and young people under the age of 18 years old.

regarding certain content[30]	Provision of mechanisms to report and make complaints regarding breaches of terms of use	Provision of accessible information on how to make reports or complaints to the eSafety Commissioner
Expectations regarding making certain information accessible	-	Provision of accessible information on terms of use, policies and complaints, etc. Provision of updates about changes in policies, terms and conditions, etc.
Record-keeping	-	Records kept regarding certain matters
Dealings with eSafety Commissioner	Provision of requested information to eSafety Commissioner	Implementation of a designated contact point

Source: Online Safety (Basic Online Safety Expectations) Determination 2022, 20 January 2022, available at: https://www.legislation.gov.au/Details/F2022L00062

4.27 Compliance with the BOSE hinges on eSafety's powers to require reporting from platforms in relation to how they are meeting expectations, which is bolstered by a civil penalties scheme in addition to other enforcement mechanisms. eSafety also has the power to publish statements in relation to how particular services are meeting expectations.[31]

Additional industry codes or standards

4.28 The OSA provides for the registration of new and improved industry codes or standards, requiring the Australian digital industry to take measures to address online safety. eSafety argued that the codes or standards will establish a 'regime of modernised industry codes or standards, expanded to additional sections of the online industry to make sure the whole digital ecosystem is playing its part'.[32]

30 The forms of content this section is directed at includes: cyber-bullying material directed at an Australian child or adult, an non-consensual intimate image of a person, class 1 and class 2 material, material promoting, inciting, instructing in or depicted abhorrent violent content.

31 eSafety Commissioner, *Submission 53*, p. 39.

32 eSafety Commissioner, *Submission 53*, p. 18.

4.29 These codes and standards are to address how online platforms can assist users to manage and limit access to harmful content, particularly in relation to Class 1 and Class 2 forms of content under the Online Content Scheme.[33]

4.30 The codes and standards are to apply to eight key forms of actors in the digital industry, which will include providers of content such as social media and other online communication services, internet and hosting providers, manufacturers and suppliers of internet-related equipment, and providers who install and maintain such equipment.[34]

4.31 A key feature of the codes and standards is that they are mandatory, with responsibility lying with eSafety in directing compliance. Non-compliance with a direction from eSafety to comply with these codes or standards may attract a civil penalty of 500 penalty points (approximately $111,000 for individuals and $555,000 for organisations).[35]

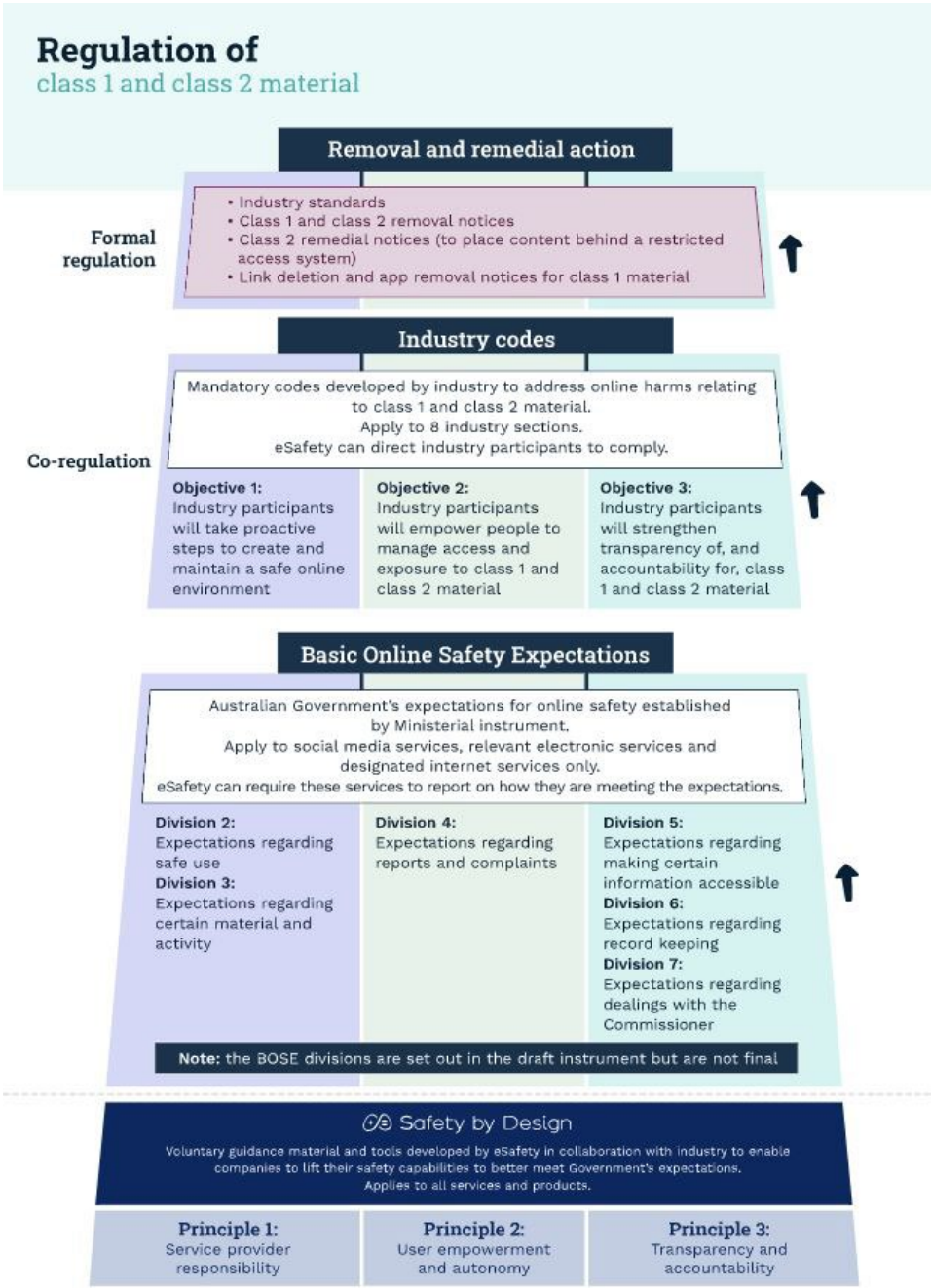
4.32 Figure 4.1 provides a breakdown of how eSafety expects the regulation of Class 1 and Class 2 material to operate. Moreover, Figure 4.1 demonstrates the shift from a self-regulatory regime towards a co-regulatory framework and – where necessary – utilising a harder formal approach.

33 Infrastructure, *Submission 44*, p. 5.

34 Infrastructure, *Submission 44*, p. 6.

35 eSafety Commissioner, *Submission 53*, p. 19.

Figure 4.1 Regulation of Class 1 and Class 2 Material under the *Online Safety Act 2021*



Source: eSafety Commissioner, Development of industry codes under the Online Safety Act – Position Paper, September 2021, available at: https://www.esafety.gov.au/sites/default/files/2021-09/eSafety%20Industry%20Codes%20Position%20Paper.pdf *(accessed 11 February 2022), p. 37.*

4.33 The eSafety Commissioner explained how these codes were developed during consultations:

> The approach that we took to these codes, just to remind you what these codes do, is that they require eight subsectors of the online industry to proactively detect and remove class 1 content and then restrict access to class 2 content. What we decided to do in consultation with the industry was to say, 'You can take the legislation and what's in the executive memorandum and come up with a set of codes and then we can decide whether we'll register them.' We tried to do some of the heavy lifting by putting four months into helping demystify and come up with a paper that had very clear outcomes and the risks and harms that we were seeking the companies to prevent, and they then had to come back to us.[36]

4.34 The eSafety Commissioner also made clear that if she was not satisfied with the codes developed by industry actors, she was not obliged to register them and could move towards the implementation of a mandatory standard to cover the industry at large.[37]

4.35 The OSA requires that 'reasonable efforts' should be made to ensure that an industry code is registered and in place within six months of the commencement of the Act (prior to 23 July 2022). If a code cannot be agreed to, an industry standard should be registered within 12 months of commencement (23 January 2023).[38] The Digital Industry Group Inc (DIGI) advised the Committee that it was drafting parts of the codes relating to social media, search engines and app distribution services.[39]

Other legislative measures

4.36 Legislative measures outside of the OSA and related regulation tend to focus primarily on law enforcement powers and counter-terrorism measures. Within the last decade, Australia's security and law enforcement agencies have gained extensive powers to investigate and disrupt criminal acts that occur online or which are facilitated by digital technologies.

4.37 Three notable recent expansions of law enforcement powers include:

---

36 Ms Julie Inman Grant, Office of the eSafety Commissioner, *Committee Hansard*, 3 February 2022, p. 17.

37 Ms Julie Inman Grant, Office of the eSafety Commissioner, *Committee Hansard*, 3 February 2022, p. 17.

38 Infrastructure, *Submission 44*, p. 5.

39 Digital Industry Group Inc (DIGI), *Submission 46*, p. 1.

- The *Surveillance Legislation Amendment (Identify and Disrupt) Act 2021* (the SLAID Act)[40];
- The *Telecommunications and Other Legislation Amendment (Assistance and Access) Act 2018* (the Assistance and Access Act)[41]; and
- The *Telecommunications Legislation Amendment (International Production Orders) Act 2021* (the IPO Act).[42]

4.38 The Assistance and Access Act grants security agencies the power to require telecommunications providers to build or use capabilities to give law enforcement targeted access to online data. Further, the Act gives law enforcement and security agencies powers to:

- Covertly access ('hack') devices;
- Search devices such as laptops, mobile phones and USBs, and collect information; and
- Conceal the fact that a device has been accessed.[43]

4.39 The SLAID Act permits Government agencies to take covert control of online accounts (for example, email or social media accounts), to 'add, copy, alter and delete data in computers', and to collect information from devices that are used, or likely to be used, by the subject of a warrant.[44]

4.40 The IPO Act is intended to improve Australian law enforcement agencies' access to data held overseas by foreign-based companies who operate in Australia. The Department of Home Affairs (argued that the IPO Act will assist in the investigation of a range of technology-dependent offences including ransomware attacks, child sexual exploitation and abuse, and serious and organised crime.[45]

---

40 *Surveillance Legislation Amendment (Identify and Disrupt) Act 2021*, available at: https://www.legislation.gov.au/Details/C2021A00098 (accessed 4 February 2022).

41 *Telecommunications and Other Legislation Amendment (Assistance and Access) Act 2018*, available at: https://www.legislation.gov.au/Details/C2018A00148 (accessed 4 February 2022).

42 *Telecommunications Legislation Amendment (International Production Orders) Act 2021*, available at: https://www.legislation.gov.au/Details/C2021A00078 (accessed 4 February 2021).

43 Department of Home Affairs (Home Affairs), *Assistance and Access: Overview*, available at: https://www.homeaffairs.gov.au/about-us/our-portfolios/national-security/lawful-access-telecommunications/assistance-and-access-overview (accessed 2 February 2022).

44 Australian Federal Police, *Fast Facts: AFP Powers under SLAID legislation*, available at: https://www.afp.gov.au/news-media/media-releases/fast-facts-afp-powers-under-slaid-legislation (accessed 2 February 2022).

45 Home Affairs, *Submission 40*, p. 6.

4.41 A number of other reforms by the Australian Government have been undertaken in the last five years which address aspects of online safety. These include:

- 'Carly's Law' criminalising acts online to prepare or plan to harm, or engage in sexual activity, with a person under 16 years of age;[46]
- New criminal offences for 'grooming' a third party and for facilitating dealings with child abuse material, as well as increased penalties and presumptive minimum sentences for child sex offences;[47]
- Legislation restricting online gambling promotion;[48]
- New civil and criminal penalties to address the non-consensual sharing of intimate images;[49]
- New offences directed at internet service providers (including hosts or content providers) for failure to report or remove live or streaming violent content, enacted in the wake of the 2019 Christchurch terrorist attack;[50] and
- The introduction of the Australian Code of Practice on Disinformation and Misinformation, which is a voluntary code which commits member companies to reduce the risk of online misinformation.[51]

Social Media (Anti-Trolling) Bill 2021 and defamation reform

4.42 The Australian Government has introduced a bill to enable Australians to 'unmask' anonymous trolls who post defamatory material online.

4.43 The *Social Media (Anti-Trolling) Bill 2022* (the Bill) is designed to address the findings of the High Court of Australia in the *Voller* decision. The Court held that owners of webpages or social media accounts can be held liable for

---

46 *Criminal Code Amendment (Protecting Minors Online) Act 2017*, available at: https://www.legislation.gov.au/Details/C2017A00050 (accessed 7 February 2022).

47 *Crimes Legislation Amendment (Sexual Crimes Against Children and Community Protection Measures) Act 2020*, available at: https://www.legislation.gov.au/Details/C2020A00070 (accessed 7 February 2022).

48 *Communications Legislation Amendment (Online Content Services and Other Measures) Act 2018*, available at: https://www.legislation.gov.au/Details/C2018A00028 (accessed 7 February 2022).

49 *Enhancing Online Safety (Non-consensual Sharing of Intimate Images) Act 2018*, available at: available at: https://www.legislation.gov.au/Details/C2018A00096 (accessed 7 February 2022).

50 *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*, available at: https://www.legislation.gov.au/Details/C2019A00038 accessed 7 February 2022.

51 DIGI, *Disinformation Code*, available at: https://digi.org.au/disinformation-code/ (accessed 8 March 2022).

defamatory third-party comments on public websites such as social media posts.[52]

4.44 The Bill seeks to clarify that social media page owners are not 'publishers' for defamatory material posted on their page by third parties, thereby protecting them from defamation liability. The Bill also provides for victims of defamation to identify anonymous users who post defamatory material by either engaging with a complaints mechanism with the provider in order to obtain the originator's contact details, or through an 'end-user information disclosure order' from a court.[53]

4.45 Concurrent with this inquiry, the Attorney-General's Department (AGD) conducted a public consultation into the exposure draft of the Bill.[54] The Bill was subsequently introduced in the House of Representatives on 10 February 2022.[55] The Senate Legal and Constitutional Affairs Legislation Committee was referred the provisions of the Bill on 10 February 2022, and is due to report its findings to the Senate on 24 March 2022.

4.46 Given this, the Committee did not enquire in depth into the proposed anti-trolling legislation. However a broad range of evidence was gathered during the course of the public hearings and submissions regarding online harm caused by trolling that may fall below the threshold of the adult cyber-abuse scheme. Such abuse includes but is not limited to, posts by anonymous trolls making damaging, defamatory remarks online.

4.47 The provisions in this Bill aim to assist victims of online, anonymous trolling with the recourse to unmask their abusers. The Bill also aims to incentivise social media platforms to provide identifying information to victims of online defamation.

4.48 As Ms Tanya Hosch, Executive General Manager of the Australian Football League stated:

---

52 *Fairfax Media Publications Pty Ltd v Voller* [2021] HCA 27.

53 Australian Parliament, *Social Media (Anti-Trolling) Bill 2022*, available at: https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=r6831 (accessed 14 February 2022).

54 Attorney-General's Department (AGD), *Social Media (Anti-Trolling) Bill*, available at: https://www.ag.gov.au/legal-system/social-media-anti-trolling-bill (accessed 11 February 2022).

55 Australian Parliament, *Social Media (Anti-Trolling) Bill 2022*, available at: https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=r6831(accessed 14 February 2022).

> I think the biggest difficulty is the fake accounts and the ability of people to set up online accounts without having to reveal their own personal identities. It means it can often be incredibly hard to trace these people. We put our reports of this material through our integrity department, which is staffed almost entirely, I think, by former police who've got significant experience in investigation work. They work closely with our social media team, and they will also reach out and liaise with the eSafety Commissioner. But that combination of effort not only is incredibly time-consuming and cumbersome but frequently leads us to a dead end because of the identities not being known to us. Even if they are members of an AFL club, we won't necessarily know that. Sometimes they will repeat the behaviour towards particular players or individuals through various identities that they set up. That is a constant frustration.[56]

4.49 In addition to the Bill, a process is underway to review the Model Defamation Provisions in state and territory laws. The reform process, which was initiated by the Council of Attorneys-General in 2018, is being led by NSW, and is currently determining how to address the question of internet intermediary liability in defamation for the publication of third-party content.[57]

Online privacy law reform

4.50 The Australian Government is currently planning to overhaul privacy law in relation to online services. It has proposed to strengthen the *Privacy Act 1988* (Privacy Act) by passing the Privacy Legislation Amendment (Enhancing Online Privacy and Other Measures) Bill 2021 (the Online Privacy Bill). According to the AGD, the Online Privacy Bill will introduce a binding online privacy code for social media and other platforms, in addition to increasing penalties for breaches and enforcement measures.[58] The Online Privacy Bill would also provide for the development of an online privacy code aimed at social media, data brokerage, and other large digital platforms. The Information Commissioner and Privacy Commissioner, Ms Angelene Falk, stated that this would 'require them to be more

---

56 Ms Tanya Hosch, Executive General Manager Inclusion and Social Policy, Australian Football League, *Committee Hansard*, 1 February 2022, p. 11.

57 NSW Department of Communities and Justice, *Review of Model Defamation Provisions*, available at: https://www.justice.nsw.gov.au/defamationreview (accessed 7 February 2022).

58 AGD, *Online Privacy Bill Exposure Draft*, available at: https://consultations.ag.gov.au/rights-and-protections/online-privacy-bill-exposure-draft/#:~:text=It%20enables%20the%20introduction%20of,closed%20on%206%20December%202021 (accessed 8 March 2022).

transparent about how they handle personal information with more stringent requirements and privacy rules for children'.[59]

4.51 The Australian Government is also conducting a Privacy Act Review (the Review), with a particular focus on privacy concerns in relation to social media and other digital platforms. This Review was opened for public consultation, and submissions on its Discussion Paper closed on 10 January 2022.[60] The Review will consider issues such as:

- The scope and application of the Act, including in relation to what constitutes 'personal information', existing exemptions, and situations where the collection, use and disclosure of personal information is permitted;
- The current Act's protections in relation to personal information and whether it provides an appropriate framework for promoting strong privacy practices;
- Potential powers for individuals to have direct rights of action to enforce privacy obligations;
- The potential introduction of a statutory tort for serious invasions of privacy; and
- Whether the notifiable data breach scheme is effective and its impact.[61]

4.52 Both the Online Safety Bill and the Review were said to be strongly focused on empowering users with oversight of what happens to their data online, particularly in relation to children's data. AGD stated that:

> With the reforms we are engaging in—the Online Privacy Bill and the broader review of the Privacy Act—a large theme through those pieces of work is providing greater transparency to individuals so that they know how their personal information is being used, and, specifically in relation to the Online Privacy Bill, what protections should be in place particularly to protect the privacy of children. There are specific protections for children in the Online Privacy Bill, including things like parental consent to the use of children's personal information by social media platforms, as well as some additional protections. In addition to social media platforms needing to consider how

---

59 Ms Angelene Falk, Information Commissioner and Privacy Commissioner Australia, Office of the Australian Information Commissioner, *Committee Hansard*, 28 January 2022, p. 15.

60 AGD, *Privacy Act Review – Discussion Paper*, available at: https://consultations.ag.gov.au/rights-and-protections/privacy-act-review-discussion-paper/ (accessed 8 March 2022).

61 AGD, *Privacy Act Review – Discussion Paper*, October 2021, available at: https://consultations.ag.gov.au/rights-and-protections/privacy-act-review-discussion-paper/user_uploads/privacy-act-review-discussion-paper.pdf (accessed 8 March 2022), p. 2.

> notice should be given to children specifically—and that might be different to how notice is given to adults—an onus needs to be put on social media platforms to consider whether the use of that child's personal information is fair and reasonable.[62]

4.53 Social media companies raised concerns that the intended reforms, including the implementation of an online privacy code before the conclusion of the Review, could result in inconsistencies in the processes. In response, Ms Falk stated that the legislative response was intended to address pressing concerns but also to indicate where wider reform would be required.[63]

Online safety policy and programs

4.54 A large part of Australia's governance of social media and online safety is overseen by eSafety, with a broad legislative framework and input from other agencies supporting its work. eSafety is provided administrative support by the Australian Communications and Media Authority (ACMA).

4.55 Policy matters primarily are the responsibility of the Department of Infrastructure, Transport, Regional Development and Communications (Infrastructure). Other agencies have responsibilities for aspects of maintaining online safety, such as the Home Affairs, the AGD, and others.

eSafety programs

4.56 eSafety has responsibility for a number of key policy tools that address online safety, such as:

- World-first schemes to assist victims in reporting online bullying and harassment, including two separate schemes for children and adults;
- A reporting tool for Australians who identify illegal and harmful content online, such as child sexual abuse, depiction or promotion of gratuitous crime and violence, and terrorism-related content;
- An online portal and reporting tool addressing image-based abuse;
- A research and education program to develop knowledge about online safety matters and provide safety awareness training to the broader community; and

---

62 Ms Julia Galluccio, Assistant Secretary, Information Law Branch, Integrity and Security Division, AGD, *Committee Hansard*, 28 January 2022, p. 38.

63 Ms Angelene Falk, Information Commissioner and Privacy Commissioner Australia, Office of the Australian Information Commissioner, *Committee Hansard*, 28 January 2022, p. 18.

- Facilitating programs that identify and provide assistance to people most at risk of online harm, including those with low digital literacy skills, Aboriginal and Torres Strait Islander peoples, and people who are not able to understand English.[64]

Online Safety Charter

4.57 The Australian Government established an Online Safety Charter (the Charter), released in December 2019, as part of the Keeping Our Children Safe Online Package. The Charter 'articulates a set of community-led expectations for industry to protect citizens, especially children and vulnerable members of the community, from harmful online experiences'.[65]

4.58 The Charter sets out several expectations of internet service providers and digital products, including the importance of safety principles, protections and processes incorporated in their design and operation.[66] The Charter is designed to apply to multiple types of digital products, including social media and networking services, content hosts, gaming providers and app developers.[67]

Additional measures across multiple portfolios

4.59 Responsibility for managing online safety is distributed across a range of Australian Government agencies. Departments, agencies and statutory bodies with a substantial online safety role include:

- Infrastructure;
- Home Affairs, including the Australian Federal Police (AFP);
- AGD;
- the Department of Foreign Affairs and Trade;
- ACMA (which includes eSafety);

---

64 Reports under eSafety's reporting schemes in relation to cyberbullying, adult cyber abuse, image-based abuse and illegal and restricted content can be made at https://www.esafety.gov.au/report/forms.

65 Infrastructure, *Online Safety Charter*, December 2019, available at: https://www.infrastructure.gov.au/media-centre/publications/online-safety-charter-0 (accessed 8 February 2022).

66 Infrastructure, *Online Safety Charter*, December 2019, available at: https://www.infrastructure.gov.au/media-centre/publications/online-safety-charter-0 (accessed 8 February 2022).

67 Infrastructure, *Online Safety Charter*, December 2019, available at: https://www.infrastructure.gov.au/sites/default/files/online-safety-charter_0.pdf (accessed 8 February 2022).

- the Australian Electoral Commission (AEC);
- the Australian Centre to Counter Child Exploitation; and
- the Office of the Australian Information Commissioner (ACCCE).

4.60 Further, some departments oversee policy or service areas that contain matters in relation to online safety, including the Department of Health, the Department of Social Services, the Department of Education, Skills and Employment, and others.

4.61 Given the fragmented oversight of different aspects of online safety, coordinating activity is necessary to achieve positive policy outcomes. Infrastructure listed a number of interdepartmental committees and working groups of which it is a part. These include:

- the Agency Heads Committee on Online Safety, an online safety focused group comprised of heads of Commonwealth departments and agencies;
- the eSafety Advisory Committee, chaired by the eSafety Commissioner and includes representatives from industry, government, civil society and academia;
- the ACCCE Prevention Awareness Working Group, which is led by the AFP;
- the Preventing Terrorist and Violent Extremist Exploitation of the Internet committee, which is led by Home Affairs; and
- the Electoral Integrity Assurance Taskforce, led by the AEC.[68]

4.62 Other measures adopted by the Australian Government include:

- An online safety measures package to boost women's safety online, including funding pilot technology to detect image-based abuse content posted online;
- Additional resourcing for frontline workers dealing with victims of technology-facilitated abuse targeting women and children; and
- The Be Connected program, which targets older Australians in improving technological literacy, build confidence in using technology and maintaining safety online.

International jurisdictions and online safety

4.63 Jurisdictions around the world have been forced to examine the issue of online safety in their legislative systems. Reset Australia provided a breakdown of the comparative legislative frameworks across the European

68 Infrastructure, *Submission 44*, p. 14.

Union, Canada, Germany, the United Kingdom (UK), Ireland and Australia, provided at Figure 4.2.

Figure 4.2 Comparative approaches to types addressing harms through regulation

	EU	Canada	Germany	UK	Ireland	Australia
Key legislation addressing harms	Digital Services Act (in draft)	Online safety proposals (in draft)	NetzDG, and others (passed)	Online Safety Bill (in draft)	Online safety & Media Regulation (in draft)	Online Safety Act (passed)
Definition of Harm, Individual, Community or Societal	No set definition, the focus is on harms that violate rights. This will include societal harms, and community harm through hate speech	Individual (aligned to existing definitions of hate speech) Societal (damage to societal cohesion, vulnerable groups)	Based on existing criminal law. This includes Individual and some community harms through hate speech	Individual (Content having an adverse physical or psychological response on adults of children)	Individual (Illegal content, individually intimidating or threatening content, eating disorder, self harm & suicide content)	Individual (content that is "offensive" to adults or children, content that is refused classification etc)
Systems Vs Takedown	Systems + Takedown	Takedown	Takedown	Systems + Takedown	Systems + Takedown	Takedown (+ potentially some systems through co-regulatory Codes)
Content In Scope	Illegal + indirectly, legal Disinfo included indirectly Hate speech indirectly included	Illegal Disinfo out of scope Hate speech in scope	Illegal Disinfo out of scope Hate speech in scope	Illegal + legal List of harms to be added later but unclear whether disinfo & hate speech is in scope (could be in scope where content is harmful to adults)	Illegal + legal Disinfo out of scope Individual hate speech content could be in scope, where it intimidates, threatens, humiliates or persecutes	Illegal + legal Disinfo out of scope Individual hate speech content could be in scope, where it causes offence to an individual or would be considered menacing, harassing or offensive
Services In Scope	Intermediary services e.g. ISPs and online platforms Private messaging out of scope	Social media Private messaging out of scope	Social media	Services which host or facilitate UGC, apart from news media outlets. Private messaging in scope.	Broad range of platforms and services inc press publications which enable UGC Private messaging in for criminal content	Social media services, Relevant electronic service and ISPs (Tight definition of "social media")
Powers Of Regulator	Fines Information gathering powers Algorithmic audit mandatory	Information gathering powers Inspection powers No algorithmic audit	Fines	Fines Information gathering powers Language seems to allow algorithmic inspection	Fines Information gathering powers. No algorithmic audit	Fines Offers public facing complaint mechanisms, Investigation, Audit (not algorithmic)
Independence Of Regulator	Independent as well as EC oversight of large platforms	Independent Creates Digital Safety Commissioner and Digital Recourse Council of Canada,	Independent	Independent however OSB keeps provisions for political agenda setting	Independent Creates Online Safety Commissioners	Independent
Transparency	Six monthly transparency reports (publicly published) Data access for pre-vetted researchers	Transparency reporting inc data on takedown volumes and processes.		Annual transparency reports No data sharing provisions	Periodic transparency reporting	Transparency reporting

Source: Reset Australia, Submission 12, pp 12-13.

The United Kingdom

4.64 The UK shares many of the same experiences and concerns with respect to online safety as Australia, and has been considering how its regulatory

environment should be updated to ensure that social media and internet companies begin to meet the community's safety expectations.

4.65 An online harms White Paper was published in April 2019 highlighting a similar range of problems to those dealt with in this inquiry. The key recommendation of the white paper was the introduction of a duty of care for internet companies, including social media and content-sharing platforms.[69]

4.66 Draft legislation was introduced in May 2021, and a Joint Committee on the Draft Online Safety Bill (the JCDOSB) was established in June of that year to scrutinise the Bill.

4.67 The JCDOSB heard that there are three key features to the UK's proposed approach to online safety:

1 It is systemic, in that platform operators are expected to have processes in place to identify, assess and mitigate hazards caused by their products

2 It is flexible, in that it applies to operators of all kinds and sizes, and

3 It is future-proof, in that it is a framework the technical details of which can be updated over time.[70]

4.68 Baroness Beeban Kidron, a member of the JCDOSB, said that the end goal of the United Kingdom's reforms was to create an online environment:

> in which the companies are actually being put under the same basic product safety approach that the rest of business is put under … [A]ny other company has to provide a product that is fit for human consumption. We do not allow cars to be put on the road with no brakes. This is a sector which is a car with no brakes.[71]

4.69 In addition to the draft Online Safety Bill, the UK has passed legislation to create an Age-Appropriate Design Code. This is a statutory code of practice covering services 'normally provided for remuneration, at a distance, by

---

69 Department for Digital, Culture, Media and Sport, *Online Harms White Paper*, available at: https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper (accessed 8 February 2022).

70 Professor Lorna Woods, Professor of Internet Law, University of Essex, *Committee Hansard*, 27 January 2022, p. 36.

71 Baroness Beeban Kidron, Peer, House of Lords, United Kingdom, *Committee Hansard*, 27 January 2022, pp 40-41.

electronic means and at the individual request of a recipient of services' which are likely to be accessed by children.

4.70 The Code includes 15 standards which should inform the design of digital services. The standards require platforms to employ child-centred design principles, and measures like data protection, age appropriateness, transparency, child-friendly defaults, and parental controls, among others.[72]

The European Union

4.71 The European Union (EU) and its shared institutions, including the European Council, European Parliament and the European Commission, have established a range of regulations in relation to online safety. As stated by the Centre for Digital Wellbeing (CDW), the EU is:

> striving to be global role model for the digital economy and internationally promote its digital standards. As such a large and influential market, their regulation of social media contributes to the setting of global norms.[73]

4.72 The EU's regulatory model is underpinned by the proposed *Digital Markets Act* and *Digital Services Act*, which will require approval from the European Council and the European Parliament prior to their implementation and subsequent binding on member states. The *Digital Markets Act* is designed to address matters such as privacy concerns, data sharing, and advertising based on personal data.[74] The *Digital Services Act* is focused on online safety matters and the protection of rights and privacy of users. The CDW provided an outline of its core proposals:

> This proposal tackles core operations of platforms, namely how information is prioritised and presented on its online interface. Significant online platforms (with more than 45 million end-users, or an equivalent of 10% of the European Union population) would be required to ensure recipients are appropriately informed of the information presented to them. The Act defines the responsibilities of digital services providers, specifically online platforms, social media, and online marketplaces. Further, it outlines obligations and procedures to tackle illegal content and disinformation, and offers the opportunity to challenge content moderation decisions. The proposal introduces safeguards protecting fundamental rights, allowing citizens to

---

72 United Kingdom Information Commissioner's Office, *Code Standards*, https://ico.org.uk/for-organisations/guide-to-data-protection/ico-codes-of-practice/age-appropriate-design-a-code-of-practice-for-online-services/code-standards (accessed 1 February 2022).

73 Centre for Digital Wellbeing (CDW), *Submission 47*, p. 39.

74 CDW, *Submission 47*, p. 40-41.

freely express themselves while maintaining rights to effective remedies, non-discrimination, the rights of the child, and personal data and privacy protection.[75]

4.73 On 20 January 2022, Members of the European Parliament agreed to the draft measures contained in the *Digital Services Act*, which is now with the European Council for approval.[76] The European Council agreed to the proposed measures in the *Digital Markets Act* on 25 November 2021.[77]

4.74 EU Member States have taken additional measures individually, including:

- Germany, which introduced the *Network Enforcement Act* in 2017 to address hate speech and misinformation online;
- Austria, which established the *Communications Platforms Act* in 2021 which is aimed at hate speech, harassment and false information on online platforms;
- Sweden, with the development of a handbook for communicators in public administration that addresses misinformation campaigns;
- Spain, which has implemented a digital transformation policy over five years in addition to adopting a digital charter of rights;
- Denmark, which has adopted a range of measures such as legislation inspired by the Australian *News Media and Digital Platforms Mandatory Bargaining Code* in relation to advertising, in addition to the adoption of the EU Code of Practice on Disinformation and increased digital literacy education for adults and children; and
- France, with the adoption of laws in relation to disinformation and misinformation, particularly during election periods.[78]

New Zealand

75 CDW, *Submission 47*, p. 41.

76 European Parliament, *Digital Services Act: regulating platforms for a safer online space for users*, Press Release, 20 January 2022, available at: https://www.europarl.europa.eu/news/en/press-room/20220114IPR21017/digital-services-act-regulating-platforms-for-a-safer-online-space-for-users (accessed 8 March 2022).

77 European Council, *Regulating 'big tech': Council agrees on enhancing competition in the digital sphere*, Press Release, 25 November 2021, available at: https://www.consilium.europa.eu/en/press/press-releases/2021/11/25/regulating-big-tech-council-agrees-on-enhancing-competition-in-the-digital-sphere/ (accessed 8 March 2022).

78 CDW, *Submission 47*, pp 46-49.

4.75 The New Zealand model of regulatory response to online safety is similar to that adopted by the Australian Government.

4.76 The primary legislative instrument regulating online safety in New Zealand is the *Harmful Digital Communications Act 2015* (HDC Act). According to the New Zealand Ministry of Justice (Tāhū o te Ture), the HDC Act is designed to 'prevent and reduce the impact of cyberbullying and other modern forms of harassment and intimidation', such as criminalising the sending of messages and/or posting material online designed to deliberately cause serious emotional distress to a victim.[79] As at 9 March 2020, the HDC Act has resulted in:

- 148 criminal charges filed relating to 115 people
- 100 criminal charges finalised relating to 79 people
- 66 convicted and sentenced
- 25 withdrawn
- 9 other outcomes, including diversions completed and dismissals.[80]

4.77 The New Zealand Parliament (Pāremata Aotearoa) recently amended the HDC Act via the Harmful Digital Communications (Unauthorised Posting of Intimate Visual Recording) Amendment Bill (the HDC Bill). The HDC Bill focuses primarily on image-based abuse and non-consensual sharing of intimate images. The HDC Bill was passed by the Parliament in March 2022 and is now enforceable.[81]

4.78 The HDC Act also established NetSafe as an 'approved agency' to assess, investigate and manage complaints in relation to online safety, in addition to the introduction of a civil court process for serious or repeated offences. As

---

79 New Zealand Ministry of Justice (Tāhū o te Ture), *Hundreds helped by cyberbullying laws*, 9 March 2020, available at: https://www.justice.govt.nz/about/news-and-media/news-and-media-archive/news-archive/hundreds-helped-by-cyberbullying-laws/ (accessed 21 February 2022).

80 New Zealand Ministry of Justice (Tāhū o te Ture), *Hundreds helped by cyberbullying laws*, 9 March 2020, available at: https://www.justice.govt.nz/about/news-and-media/news-and-media-archive/news-archive/hundreds-helped-by-cyberbullying-laws/ (accessed 21 February 2022).

81 New Zealand Parliament (Pāremata Aotearoa), *Harmful Digital Communications (Unauthorised Posting of Intimate Visual Recording) Amendment Bill*, available at: https://www.parliament.nz/en/pb/bills-and-laws/bills-proposed-laws/document/BILL_99360/harmful-digital-communications-unauthorised-posting-of (accessed 21 February 2022).

at 9 March 2020, the courts had received 14 civil cases requesting Harmful Digital Communications Orders, six of which were completed.[82]

4.79 NetSafe, a similar body to Australia's Office of the eSafety Commissioner, is a not-for-profit independent organisation, and works with the New Zealand Government in relation to online safety, including education and research. NetSafe also operates a reporting service, where people can report matters such as fraud, privacy issues, and online bullying or harassment.[83] NetSafe and the New Zealand Ministry of Justice also provide information for parents and organisations (such as schools) on online safety guidelines and strategies. The New Zealand Department of Internal Affairs (Te Tari Taiwhenua) also has responsibility for matters in relation to online child exploitation, governed by the *Films, Videos and Publications Classification Act 1993*.[84]

4.80 NetSafe is currently leading the development of the Aotearoa New Zealand Code of Practice for Online Safety and Harms, the voluntary industry code to establish a self-regulatory framework for the digital industry and agree on certain principles and commitments. Companies such as Meta, Google, Microsoft, TikTok, Twitch and Twitter have been involved in the drafting of the new code, which as at 22 February 2022 was open to public comment.[85]

United States of America

4.81 The United States of America (the US, the United States) has a number of legislative protections, particularly in relation to privacy. The following federal legislation is in place which governs aspects of online safety:

- The *Electronic Communications Privacy Act*, enabling the US Government to access digital communications and tracking technology with a subpoena;

---

82 New Zealand Ministry of Justice (Tāhū o te Ture), *Hundreds helped by cyberbullying laws*, 9 March 2020, available at: https://www.justice.govt.nz/about/news-and-media/news-and-media-archive/news-archive/hundreds-helped-by-cyberbullying-laws/ (accessed 21 February 2022).

83 NetSafe New Zealand, *Report an Incident*, available at: https://www.netsafe.org.nz/reportanincident/ (accessed 21 February 2022).

84 New Zealand Department of Internal Affairs (Te Tari Taiwhenua), *Online Digital Child Exploitation*, 2021, available at: https://www.dia.govt.nz/digital-child-exploitation (accessed 22 February 2022).

85 NetSafe New Zealand, *Aotearoa New Zealand Code of Practice for Online Safety and Harms Draft*, 2 December 2021, available at: https://www.netsafe.org.nz/aotearoa-new-zealand-code-of-practice-for-online-safety-and-harms-draft/ (accessed 21 February 2022).

- The *Computer Fraud and Abuse Act*, which criminalised the access and sharing of protected information;
- The *Communications Decency Act*, which provides extensive immunity from liability for publishers and users of online platforms; and
- The *Children's Online Privacy Protection Act*, in effect from 1998, and its associated Rule, requiring that websites collecting information in relation to children under the age of 13 years of age to comply with the Federal Trade Commission.[86]

4.82 Importantly, as the Australian Law Reform Commission has noted, there are no privacy protections at federal law aimed at adults in the United States.[87] Restrictions on harmful content are limited in United States legislation due to the limitations placed on such attempts by the First Amendment of the US Constitution. Multiple Supreme Court cases have struck down regulation in favour of less restrictive methods of moderation (such as utilising filtering technology).[88]

4.83 Further regulation nonetheless continues to be pursued. There have been a number of Congressional and Senate hearings in recent years, including the appearance of Ms Frances Haugen at a hearing by the Senate Committee on Commerce, Science and Transportation.[89] President Joe Biden has flagged online safety for children as a key topic of concern for his administration, and used the 2022 State of the Union address to request action on these matters.[90]

---

86 CDW, *Submission 47*, pp 33-34.

87 Australian Law Reform Commission, *For Your Information: Australian Privacy Law and Practice,* ALRC Report 108, August 2010, available at: https://www.alrc.gov.au/publication/for-your-information-australian-privacy-law-and-practice-alrc-report-108/69-particular-privacy-issues-affecting-children-and-young-people/online-consumers-and-direct-marketing-issues/ (accessed 23 February 2022).

88 Internet New Zealand, *Regulatory tools to address harm from content and conduct online*, June 2020, available at: https://internetnz.nz/assets/Archives/Regulatory-tools-to-address-harms-from-content-and-conduct-online.pdf (accessed 23 February 2022), p. 38.

89 CDW, *Submission 47*, p. 33.

90 White House, *FACT SHEET: President Biden to Announce Strategy to Address Our National Mental Health Crisis, As Part of Unity Agenda in his First State of the Union*, 1 March 2022, available at: https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/01/fact-sheet-president-biden-to-announce-strategy-to-address-our-national-mental-health-crisis-as-part-of-unity-agenda-in-his-first-state-of-the-union/ (accessed 4 March 2022).

Further change to consolidate legislative powers

4.84 As detailed above, the legislative and policy responsibility for matters relating to online safety are divided amongst multiple portfolios with varying functions, resulting in powers being dispersed broadly. Some submitters argued that the OSA should be reviewed in future to ascertain whether it had made an impact in relation to online safety and to examine whether a single regulatory framework would reduce complexity and confusion for providers.

4.85 Digital Industry Group Inc. (DIGI) commented on the disparate nature of legislation pertaining to online safety, suggesting that a single regulatory framework would be less confusing for providers. An example of this trend was provided in relation to abhorrent content, which DIGI stated is covered by multiple schemes in the OSA, but is also captured by the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*. This, DIGI argued, has led to the emergence of inconsistencies across the legislative instruments:

> As one example of an inconsistency, the OSA's takedown schemes and the BOSE suggest that service providers should be required to remove all types of Class 1 material. However, the Commissioner's position as stated in their position paper on the OSA Codes is that an identified subclass of Class 1, termed "Class 1b (fetish practices)" can be treated as Class 2 materials, and therefore do not need to be removed. It is unclear whether this interpretation extends to other aspects of the OSA, which creates confusion for industry participants working in good faith to comply with the legislation.[91]

4.86 DIGI also noted that other jurisdictions, such as the European Union and the UK, were seeking to implement an overarching legislative framework. It argued that a similar framework in Australia would:

> aid clarity and compliance – particularly for start-ups, smaller challenger companies, and those without a large local staff presence – who may be struggling to make sense of the complex regulatory environment in Australia.[92]

---

91 DIGI, *Submission 46*, p. 4.

92 DIGI, *Submission 46*, p. 4.

4.87 While DIGI acknowledged that the OSA improves this situation, it recommended that the Australian Government adopt a single consolidated legislative framework to ensure clarity and certainty for providers.[93]

4.88 Many industry members and bodies pointed out that the Australian Government had passed a large number of legislative amendments and conducted a similarly large number of reform consultation processes. For instance, Meta stated that the Australian Government had already heavily regulated the online space in recent years, and it was unclear to what extent these reforms were having an effect:

> In the last three years, at least 14 new federal regulations have come into force which primarily impact digital platforms. There have also been at least 18 major Government or parliamentary inquiries or consultations impacting digital platforms over the last three years. These developments are on top of existing regulations that cover digital platforms, including online safety, privacy and multinational taxation laws. The Government has recently foreshadowed additional regulations, including digital platforms-specific competition laws; age or identity verification; and new obligations about working with law enforcement.[94]

4.89 Meta argued that attention should therefore be focused primarily on whether the recent changes were working effectively, rather than examining whether further ones were required.[95] It also noted the risk of 'overlapping, duplicative or inconsistent rules across different laws', arguing that this hinders platforms in effectively understanding and implementing the legislation into its systems.[96]

4.90 The global and interconnected nature of the digital world was also highlighted by Meta, who suggested that an Australian regulatory path should be cognisant of the 'global contest of competing visions of the internet'.[97] Meta argued that other nations watch Australia's regulatory approach in mind of their own legislative approaches, and that regulators should examine 'whether Australian regulation sets an example which

---

93 DIGI, *Submission 46*, p. 4.

94 Meta, *Submission 49*, p. 4.

95 Meta, *Submission 49*, p. 4.

96 Meta, *Submission 49*, p. 5.

97 Meta, *Submission 49*, p. 5.

encourages a liberal, open and democratic approach to the internet, or an internet that is more closed, tightly controlled and fragmented'.[98]

4.91 Snap Inc. (Snap) were also supportive of a single regulatory framework in the model of the proposed *Digital Services Act* in the EU. It stated:

> Online regulation is most effective when it is based on broad principles that companies of all sizes are able to follow and implement proportionately, as relevant to their service and risk profile. Such regulation focuses on the principles or outcomes companies should deliver, setting out "what" objectives are to be achieved, without being too prescriptive as to "how" companies should achieve them. There is incredible variety in the size, resources and service models of different online platforms. A principles-based approach accommodates this variety and allows for innovative, effective approaches to be developed, while focusing on what is most important: the safety of users.[99]

4.92 Snap raised concerns that regulation of social media platforms could result in highly detailed and complex requirements, which only the biggest social media companies would be able to comply with due to their capacity to have large compliance teams.[100] It argued that this point had been raised by the Australian Competition and Consumer Commission, which had stated in its 2019 Digital Platforms Inquiry that large dominant social media firms stifled competition in the market and had negative impacts on consumers.[101] Snap stated that further 'overly prescription' regulation on smaller entities in the market would only amplify these power dynamics and the strength of the major social media companies.[102]

4.93 This point was also noted by the larger social media companies. Twitter argued that further regulatory intervention would 'undermine competition and entrench incumbent services, reducing consumer choice':

> Policymakers should avoid mandating technical means of implementation that have the effect of further entrenching services based on those tools and technologies, or of benefiting those that have the financial and technical means to deploy the particular implementation proposed, not to mention the vendors promising a simple solution. Opportunities to expand interoperability and the

98 Meta, *Submission 49*, p. 5.

99 Snap Inc., *Submission 16*, p. 5.

100 Snap Inc., *Submission 16*, p. 5.

101 Snap Inc., *Submission 16*, p. 5.

102 Snap Inc., *Submission 16*, p. 5.

> adoption of open standards will empower people with greater choice and flexibility about how they interact with online services and drive competition.[103]

4.94 Google further noted that the disparate nature of Australia's regulatory framework created inconsistencies in implementation, which was challenging for the industry to manage:

> For example, age verification is proposed within five of these different workstreams with varying timetables for implementation. The lack of a current, clear, evidence-based pathway for delivery makes it more challenging for the industry. A coordinated approach to these issues that is unified under a whole of Government approach would help to ensure consistency and efficiency as we work together to further improve the online well-being of all Australians.[104]

Voluntary v. mandatory requirements

4.95 A common concern raised by many witnesses was that a number of Australia's regulatory approaches were based on voluntary Codes of Practice or Codes of Conduct. Dr Michael Salter expressed the view that any form of self-regulation by the digital industry would not improve online harms:

> Whenever these scandals are exposed in the press, we hear the same reassuring platitudes from social media executives, and, when pushed, they gradually adjust their operational models to placate public and political outrage, but their responses are consistently reactive and lacking in transparency and accountability … The lessons of the last 20 years suggest that self-regulation, co-regulation and voluntary industry codes of practice are insufficient to keep children safe.[105]

4.96 This point was echoed by Dr Hany Farid, who stated his view that the technology industry is 'simply incapable of self-regulation'.[106]

---

103 Twitter, *Submission 50*, p. 6.

104 Google Australia, *Submission 30*, pp 2-3.

105 Dr Michael Salter, *Committee Hansard,* 18 January 2022, p. 10.

106 Dr Hany Farid, *Committee Hansard,* 28 January 2022, p. 2.

4.97 Witnesses pointed to examples of industry self- or co-regulation which indicate that this model has been problematic. For example, Reset Australia explained that the voluntary code of conduct in relation to misinformation and disinformation had very few requirements on social media platforms:

> If you look at how that's currently operating, there's a voluntary code that's opt in; the transparency reporting doesn't have clear KPIs or metrics; the committee charged with that is meeting on a six-monthly basis; and there are no clear compliance mechanisms or penalties in place.[107]

4.98 Ms Frances Haugen expressed disbelief at the prospect that social media and digital platforms would comply with a voluntary regulatory framework. She stated:

> [I]n every part of the world, what has become clear over the last few years is that self-regulation does not work. Platforms cannot be trusted to act in the public interest. They are often, as my revelations showed, fully aware of the harms caused by their products and services, and yet choose to ignore these in favour of growth and profit.[108]

4.99 Further, Ms Haugen stated that social media companies were unlikely to comply with any voluntary requirement which could potentially hurt their business model primarily because they are conscious of the power they hold. She explained that companies such as Meta are aware that the only people in the world who genuinely understood its systems were its employees, and it could therefore be selective about explaining how its systems work to the broader community. This, she argued, led to companies intentionally misleading the public in order to provide the appearance of compliance.[109]

4.100 The ACMA explained that voluntary regulation was designed to 'encourage and incentivise those digital platforms to take actions themselves', and that many online companies had taken steps to addressing commitments set out in voluntary codes such as in relation to disinformation and misinformation.[110] The voluntary code in relation to disinformation and

---

107 Ms Dhakshayani Sooriyakumaran, Director of Tech Policy, Reset Australia, *Committee Hansard*, 20 January 2022, p. 52.

108 Ms Frances Haugen, *Committee Hansard*, 3 February 2022, p. 8.

109 Ms Frances Haugen, *Committee Hansard*, 3 February 2022, p. 8.

110 Ms Nerida O'Loughlin, Chair and Agency Head, Australian Communications and Media Authority (ACMA), *Committee Hansard*, 20 January 2022, p. 59.

misinformation was framed as a 'first instance' response by government, and that further action may be required.[111]

Committee comment

4.101 Australia has made significant additions to its online safety regulatory regime in recent years, particularly by the introduction of its world-first adult online abuse and cyberbullying reporting schemes, as well as through the creation of the BOSE.

4.102 In addition, security and law enforcement agencies' powers to identify, locate and disrupt the activities of online abusers – particularly the sexual exploitation and abuse of children online – have been greatly enhanced. The addition of powers to covertly access and control all kinds of electronic devices and online accounts, and access information held by overseas-based technology and social media companies, represents a substantial improvement to law enforcement's ability to protect Australians from online harms.

4.103 While the commencement of the OSA early in 2022 has expanded the role played by the eSafety Commissioner and cemented the Commissioner's role as a key regulator in the online safety space, there is scope to build on that success and expand the scope of the OSA to simplify regulation.

4.104 Given the broad suite of issues that fall under the rubric of online safety, further centralisation of responsibility for online safety policy or enforcement may be challenging, unsuitable and impractical. Nonetheless, the Committee is mindful of the evidence adduced by industry groups and members who pointed to examples of inconsistency and uncertainty.

4.105 The OSA is required to have an independent review in January 2025. While this will enable the effects of the OSA to be seen more fully and in context with the other legislative reforms and policy measures being put in place, there remains areas of the law where attention is urgently required to prevent otherwise avoidable harm.

4.106 Delaying a subsequent review of the framework may hinder further action being taken in relation to matters which currently are not fully covered by legislation, such as technology-facilitated abuse in family and domestic violence contexts, volumetric attacks, and the risks posed by encryption.

---

111 Ms Nerida O'Loughlin, ACMA, *Committee Hansard*, 20 January 2022, p. 58.

4.107 Further, a broad-scale review of digital safety more generally has never been undertaken. This would pose a considerable challenge in terms of scope, conduct and investigations. However, the Committee also believes that the need for industry to fully understand its roles and obligations while providing safety for users is necessary in a world where digital technology is fast becoming the primary form of communication for individuals, groups, business and government.

4.108 A broad review capturing all elements of digital legislative frameworks and needs for an online Australia is necessary to fully comprehend the online framework. This review should be commenced within 18 months of the OSA's establishment, to provide time to allow the effects of the legislation to be seen while ensuring that urgent action is prioritised.

4.109 Additionally, the Committee also agrees with the views put by many witnesses that the social media and digital industry has for too long been able to self-regulate, with very few results to demonstrate its success. A broad review should cover all self- and co-regulatory frameworks currently in place, in addition to voluntary codes, and assess the adequacy of compliance with these models by social media and digital services. In the event that they are determined to be insufficient, the Committee recommends that future regulatory measures be mandatory.

Recommendation 18

4.110 The Committee recommends that the Department of Infrastructure, Transport, Regional Development and Communications conduct a Digital Safety Review on the legislative framework and regulation in relation to the digital industry. The Digital Safety Review should commence no later than 18 months after the commencement of the *Online Safety Act 2021*, and provide its findings to Parliament within twelve (12) months.

Recommendation 19

4.111 The Committee recommends that, subject to Recommendation 18, the Digital Review examine the need and possible models for a single regulatory framework under the Online Safety Act, to simplify regulatory arrangements.

5. Online Safety 2.0

The Pursuit of a New Digital Culture

Introduction

5.1 As outlined in the previous chapters, online spaces present unique and difficult challenges in regulation. The digital world is increasingly vast, complex and ever-changing. Nonetheless, it is imperative on industry, governments, and private citizens worldwide to continue to meet the challenge of ensuring safety and security in online environments.

5.2 The status quo is neither desirable nor sustainable. Australia, alongside the rest of the world, must address the causes and amplifiers of harm to ensure a safe and equitable internet, especially to the most vulnerable members of society.

5.3 Notwithstanding these principles, it is clear that while the technology industry and government both have roles to play in addressing online harm, the broader society also have a critical role in ensuring online environments are safe. The Committee found during the inquiry that this third element is critical to any success of online safety programs. Without broader societal and cultural change to address these issues, any action taken by government or industry will be moot.

5.4 This chapter outlines the issues facing cultural and societal potential methods of addressing online harm. It examines how regulators can balance the freedom of expression online with the need to prevent harm and provide safety to users. It then examines potential methods of protecting users

online, such as a statutory duty of care applied to online platforms. It also considers the need for increased education about online safety, examining what digital education providers currently exist and what gaps have been identified. The report concludes with the Committee's comments and recommendations.

Balancing freedom online with harm prevention

5.5 In attempting to find solutions to address online harm, it is critically important to recognise that any regulation in this area will necessarily curtail the operations of the online world, and indeed affect the broader freedom of expression. Maintaining an appropriate balance is therefore a delicate and complex task.

5.6 The evidence presented to the Committee ultimately fell into two distinct schools of thought. One perspective suggested that a heavy emphasis on online safety had the potential of impacting freedom of speech and expression on digital platforms. The alternative perspective suggested that current models of co- and self-regulation performed by the digital industry are failing to protect vulnerable users from harms such as online abuse and child sexual abuse material.

Protecting freedom of speech

5.7 Some social media and digital platform providers indicated their concern that the regulation of online content would stymie users' capacity to express themselves and would thereby reduce personal liberty.

5.8 Among the defenders of this principle, Twitter stated that it 'recognises the need to balance tackling harm with protecting a free and secure Open Internet.'[1] Twitter explained that, while it supported regulation to the extent that it would empower people who otherwise would not speak out due to fears of abuse (which it stated was done via its Rules), it believed that a balance needed to be struck between regulation and free expression.[2] This sentiment was also expressed by Meta, who stated that it particularly considered the need to provide a balanced approached when it came to issues such as age verification.[3]

---

1 Twitter, *Submission 50*, p. 3.

2 Twitter, *Submission 50*, p. 11.

3 Meta, *Submission 49*, p. 30.

5.9 Twitter opined that the balance between harm reduction and freedom of expression was vital in discussions regarding online safety, stating that over-regulation could potentially lead to the unintended silencing of debate, particularly for marginalised groups.[4]

5.10 Twitter emphasised the need for governments to protect a 'free and secure Open Internet', which included ensuring equitable access and compliance with human rights norms such as freedom of expression:

> Governments should prioritise policies, partnerships, and investments at home and abroad that support and defend the Open Internet, both through regulatory and standards bodies, as well as ensuring domestic regulation does not undermine global norms or set dangerous precedents. Open standards championed by these bodies will provide for greater interoperability, connection, and competition.[5]

5.11 Further, Twitter argued that online harm is a reflection of offline societal harm, and that increased content moderation would not make these issues disappear:

> More broadly, the policy issues addressed are often rooted in complex societal challenges that exist in offline, as well as online, contexts. As we continue to work together in good faith on these complex issues, we emphasise that these challenges will not be resolved by the removal of content online alone. Bad actors seeking to exploit online services to undermine elections, spread disinformation, and harm others will not be deterred by their accounts being removed. Effective solutions demand a whole of society response that recognises the full scope of the problem being addressed.[6]

Improving online culture

5.12 Evidence to the Committee suggested that online culture is extremely toxic and lacks basic standards of civility and social decency. Witnesses suggested that online culture is encouraging harmful standards of behaviour for broader society. Dr Kate Hall, Head of Mental Health and Wellbeing for the Australian Football League (AFL), stated:

> As a psychologist, I think we're seeing social norms move further and further online, away from what is appropriate and socially normative in face-to-face. I

---

4 Ms Kathleen Reen, Senior Director of Public Policy, Asia-Pacific, Twitter, *Committee Hansard*, 21 January 2022, p. 21.

5 Twitter, *Submission 50*, pp 3-4.

6 Twitter, *Submission 50*, p. 3.

> think there are grave harms occurring, particularly for the mental health and wellbeing of our young people, as there is this progression away from what is socially normative behaviour in our schools, in our clubs, on the street and in public settings. Even in our stadiums these behaviours would never ever be acceptable, and any perpetrator would be addressed by all others sitting around them or walking past them; that nudges the socially normative behaviour back to a far safer and healthier environment. It doesn't mean every outlier then adheres to it, but the collective as a whole creates these expectations of those who are on that platform.[7]

5.13 One submitter stated that the nature of the social media industry's business model encourages the proliferation of toxicity. They pointed out that the extreme and toxic content is profitable to platforms, who use technological devices such as algorithms promoting extreme content and sensationalism to ensure user engagement is maximised. This in turn fuels addiction-driven cognitive responses by users:

> Platforms encourage addictive behaviour through positive intermittent reinforcement, such as limitless feed scrolling, 'like' and 'share' buttons and comment functions, creating an addictive buzz similar to poker machines. As Facebook's founding president explained, the platform intends to 'consume as much of your time and conscious attention as possible': every interaction gives the user 'a little dopamine hit' to encourage addiction. Outrage and negativity equal more engagement, which means more dopamine rewarding potentially poor behaviour. Posting something abusive or defamatory thus acquires a seductive pull: the more extreme content gets more engagement, which humans are wired to crave.[8]

5.14 Further, the submitter argued that social media 'supercharges polarisation and tribalism', further amplifying toxicity.[9] When drawn towards a particular tribe that a person identifies with, it can 'encourage group attacks, reinforcing tribal connection'.[10] It was also asserted that the forms of online abuse that proliferate on social media sites would be unlikely to occur offline:

> Social media 'pile-ons' can be devastating for the target, and the notifications of abuse are non-stop, direct to your phone. Such bullying would probably not

---

7 Dr Kate Hall, Head of Mental Health and Wellbeing, Australian Football League (AFL), *Committee Hansard*, 1 February 2022, p. 13.

8 Name Withheld, *Submission 60*, p. 17.

9 Name Withheld, *Submission 60*, pp 17-18.

10 Name Withheld, *Submission 60*, p. 18.

> occur in person. But online, we have fewer physical and visual cues to encourage empathy. Some argue social media has facilitated bullying of marginalised populations. Profiles on dating sites now regularly proclaim 'No Indians!' or 'No Asians!' – prejudices most people would not announce at a public bar, but happily broadcast online. Anonymity can further embolden abuse, hate speech and intolerance.[11]

5.15 Other witnesses suggested that the nature and extent of harm caused by online abuse was not broadly understood in the community. Ms Tanya Hosch of the AFL put the view that her interactions in her professional and personal capacities indicated that in general people are unaware of the extent of harm or duration of the harm caused by bullying and discrimination even in offline environments.[12]

5.16 Professor Amanda Third suggested that an important element of any attempt to address online harm was to focus on championing diversity and creating strong online communities:

> It's very tempting to think of these issues in a really small way, in a way that focuses on individuals and how we teach individuals to behave online. This might be one strategy that we deploy, but we also need to think about how we create vibrant communities that have cultures of acceptance, understand diversity and are sympathetic to it, and can relate to other people as human beings. We need to do much more there to think more holistically about the ways we address these kinds of issues.[13]

Industry's limited emphasis on harm prevention

5.17 The alternative view presented to the Committee was that current regulatory models, which were said to prioritise freedom of speech and enable digital platforms to co- or self-regulate their activities, have so far failed to protect users from harm.

5.18 eSafety expressed the view that the current regulatory system has placed too great an emphasis on users taking responsibility for their personal safety online:

---

11 Name Withheld, *Submission 60*, p. 18.

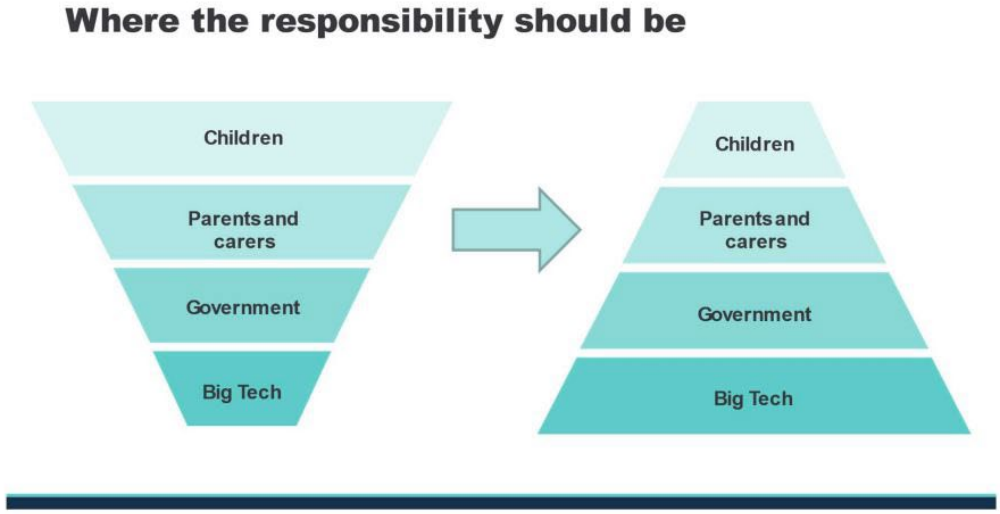
12 Ms Tanya Hosch, Executive General Manager Inclusion and Social Policy, AFL, *Committee Hansard*, 1 February 2022, p. 17.

13 Amanda Third, Professorial Research Fellow, Institute for Culture and Society, Western Sydney University; Co-Director, Young and Resilient Research Centre, Western Sydney University, *Committee Hansard*, 21 December 2021, p. 30.

> The burden of responsibility needs to be flipped so that large tech companies take responsibility for online safety and embed safety features in the design and development of their products. It should not be left to children, parents and members of vulnerable communities who experience online abuse at rates much greater than the general population to protect themselves online from harms that are enabled by the design of services.[14]

5.19 eSafety outlined this in a graph, which is provided at Figure 5.1. This figure reflects eSafety's view that responsibility should be shifted away from children and their families, towards digital platforms and government regulators.

Figure 5.1 eSafety's proposed model of the burden of responsibility



Source: eSafety Commissioner, Submission 53, p. 62.

5.20 Many witnesses echoed the concerns of eSafety, similarly perceiving an imbalance in responsibilities in keeping people safe online. Witnesses such as The Alannah and Madeleine Foundation (AMF) suggested that regulatory approaches broadly had focused primarily on educating users on ways to remain safe, rather than placing the burden of ensuring safety on social media services.[15] The AMF suggested this was due to the absence of a safety

14 eSafety Commissioner, *Submission 53*, p. 61.

15 Ms Sarah Davies, Chief Executive Officer, Alannah and Madeline Foundation (AMF), *Committee Hansard*, 21 December 2021, p. 18.

by design principle when initially designing digital platforms.[16] The AMF stated:

> An analogy is that, if you were going into a toy shop to buy a toy for your child, you would know that whatever you purchased had gone through some kind of quality control assessment. It was age appropriate. There would be no button batteries for young children or toddlers. There'd be no lead paint. You know that the system has actually put a safety-by-design element around that. That does not exist in the online world, and it should.[17]

5.21 This analogy was also used by Dr Michael Salter, who noted that most people have basic safety expectations of toys being sold, and do not 'spend our time educating parents on the pros and cons of every single toy on the shelf'.[18] Dr Salter also pointed out that, like toys, digital platforms are being produced in enormous scale, and that it is not feasible to expect parents to be aware of safety considerations for every individual platform.[19]

5.22 Basic safety expectations were highlighted as a critical component of adult online safety as well. Dr Kate Hall, Head of Mental Health and Wellbeing at the AFL, expressed concerns that victims of online abuse are expected to take on the burden of reporting and seeking resolution for online abuse:

> [T]he onus is still on the victim to have the psychological capacity, readiness and energy, and trust in the system, to meet any kind of resolution. That's not a very victim-centric or survivor-centric mindset. It doesn't empower people, particularly in our industry, to act, even though they're not bystanders; they do stand up and support each other.[20]

5.23 As discussed in Chapter 4, the eSafety Commissioner has created the voluntary Basic Online Safety Expectations (BOSE) framework as part of the *Online Safety Act 2021* (OSA) implementation. However, concerns were raised by submitters that the lack of a compulsory and enforceable duty to protect users will not encourage digital platforms to adequately protect its users.[21] Further, it is not clear that breaches of the BOSE and subsequent

---

16 Ms Sarah Davies, AMF, *Committee Hansard,* 21 December 2021, p. 18.

17 Ms Sarah Davies, AMF, *Committee Hansard,* 21 December 2021, p. 18.

18 Dr Michael Salter, *Committee Hansard,* 18 January 2022, p. 16.

19 Dr Michael Salter, *Committee Hansard,* 18 January 2022, p. 16.

20 Dr Kate Hall, AFL, *Committee Hansard,* 1 February 2022, p. 12.

21 Ms Frances Haugen, *Committee Hansard,* 3 February 2022, p. 1.

response by the eSafety Commissioner would be sufficient to prompt change.

A statutory duty of care

5.24 A number of submitters were in favour of a duty of care placed on social media platforms and other digital services, which was legally enforceable and compulsory for all digital participants.

5.25 Dr Michael Salter stated that a duty of care should be imposed on social media platforms, particularly in relation to children:

> We need social media companies to accept that they have a special duty of care to children on their platforms. We should also be cognisant of the broader impacts of social media and online platforms on sex offending as a whole. Online networks and infrastructure are awash with child sexual abuse material because technology companies are not legally obliged to proactively detect or remove that material. We have online communities of child abusers numbering in the millions taking advantage of anonymity and encryption. Establishing a baseline of online safety for children is an important step forward, but it can't be the final word, and there is a lot more work to do to ensure that social media and the internet are functioning in the best interests of children.[22]

5.26 The Centre for Digital Wellbeing (the CDW) also recommended that the Australian Government implement a regulated duty of care, suggesting that it be attached to a licencing scheme for social media platforms.[23] In support of this proposal, the CDW suggested that:

> Over time, this would likely impact on design features and algorithmic functions that are harmful to all users, but particularly youth. A clear legislative requirement such as this would compel social media companies to act on evidence that any of its features are significantly damaging.[24]

The UK model of a statutory duty of care

5.27 The UK Government has proposed the establishment of a statutory duty of care in its new bill in response to online harm.

---

22 Dr Michael Salter, *Committee Hansard,* 18 January 2022, p. 11.

23 Centre for Digital Wellbeing (CDW), *Submission 47*, p. 5.

24 CDW, *Submission 47*, p. 5.

5.28 The CDW noted that the *Online Harms White Paper* produced by the UK Government in April 2019 suggested a singular regulatory framework underpinned by a duty of care on digital platforms.[25] The subsequent development of the draft *Online Safety Bill* placed duties of care on digital companies providing content-sharing services and search functions to mitigate against illegal content.[26]

5.29 Professor Lorna Woods stated that during the development of the UK's legislative response to online safety, the Carnegie Trust had recommended a statutory duty of care, which would be supported by an independent regulator to monitor compliance and take enforcement action as required.[27] Following on from this point, Baroness Beeban Kidron OBE, a member of the Joint Committee on the Draft Online Safety Bill (JCDOSB), argued that this approach enabled digital platforms to take a flexible approach in protecting users online and develop varying approaches to online safety dependent on the functions and goals of the platform in question.[28]

5.30 Mr Damian Collins MP, Chair of the JCDOSB, explained that the role of the regulator was crucial in enforcement action on social media platforms in particular. He noted that, when drafting the JCDOSB's report, it was important to link offences in the proposed online safety regime to existing offences in law. In doing so, Mr Collins stated that the JSCOSB identified that many of the linking offences in other areas of law were drafted having never envisaged their use in a digital setting, or that the offences would be hosted by a third party such as a digital platform. He stated that this led the JSCOSB to consider the adoption of thresholds for offences online:

> The question then is: how do you introduce the thresholds? Again, going back to racial abuse, UK courts have given rulings on racial abuse. One of the cases that we looked at and discussed quite a bit in the committee was the racial abuse directed towards England footballers after the final of the European Championship. Some people have been prosecuted for what they posted, sometime after the event, so the law has already created a threshold for understanding when we think abuse has taken place, and the question then for an online safety regime would be for the regulator, through its codes of

---

25 CDW, *Submission 47*, p. 51.

26 CDW, *Submission 47*, p. 51.

27 Professor Lorna Woods, Professor of Internet Law, University of Essex, *Committee Hansard,* 27 January 2022, p. 34.

28 Baroness Beeban Kidron OBE, Peer, House of Lords, United Kingdom, *Committee Hansard,* 27 January 2022, p. 36.

> practices, to set a bar that says to the online platforms, 'This is where we believe an offence is being committed; this is where we believe you should intervene and mitigate the content that's been posted which will clearly cause harm, based on guidance that already exists in law.' The job of the regulator through the codes of practices would be to say, 'Here are offences in law that already exist and that the UK parliament has already determined are offences; these are the thresholds where we expect you to mitigate and act on it; and here are the sanctions we could apply if you fail to do so.'[29]

The best interests of the child as a guiding principle

5.31 Children's rights were suggested as a focus point for any potential reform in online safety. Ms Anne Hollands, National Children's Commissioner, Australian Human Rights Commission, argued that the child's best interests principle should be at the forefront of all business in the digital industry:

> The best interests of children should be the priority requirement for all internet based businesses. This would include strong default privacy settings and human rights by design requirements. There should be a requirement to comply with children's rights principles, such as demonstrated under the National Principles for Child Safe Organisations in the physical world. This would include a requirement to assess and report the impact on the rights of children at every stage of design, implementation and operation.[30]

5.32 The National Children's Commissioner stated that social media and other digital platforms should be required to demonstrate that their services meet the best interests of the child principle.[31] This would include considerations of privacy, security of personal data, protection from harm, a voice to express their views, and the ability to seek, receive and convey information.[32] The National Children's Commissioner further advocated that:

> A best interests approach may require implementing clear boundaries to prevent practices that both infringe upon children's rights and are contrary to their best interests, including by curtailing routine and indiscriminate digital

---

29 Mr Damian Collins MP, Chair, Joint Committee on the Draft Online Safety Bill, House of Commons, United Kingdom, *Committee Hansard,* 27 January 2022, p. 37.

30 Ms Anne Hollands, National Children's Commissioner, Australian Human Rights Commission (AHRC), *Committee Hansard*, 2 March 2022, pp 1-2.

31 Ms Anne Hollands, National Children's Commissioner, AHRC, *Committee Hansard*, 2 March 2022, p. 3.

32 National Children's Commissioner, *Submission 64*, p. 4.

> surveillance measures. Practices such as online tracking, profiling, behavioural monitoring and 'nudging', the collection of biometric and geolocation data from children, automated decisions affecting children and the unjustifiable sale or transfer of children's personal data to third parties should be banned or heavily restricted to protect children's rights.[33]

5.33 Similarly, the AMF recommended that the 'best interests of the child' principle be adopted in considering changes to regulation of online safety, stating that 'If we do that, then all the steps that we take to uphold children's best interests on digital platforms will be as constructive and healthy as possible'.[34] Other organisations that were supportive of the implementation of the best interests of the child principle included ReachOut[35] and Orygen.[36]

5.34 The AMF also suggested that any changes uphold the rights contained in the United Nations Declaration of Children's Rights, including the right to privacy, safety, dignity and expression.[37] The National Children's Commissioner also pointed to other international agreements outlining the rights of children, including the Convention on the Rights of the Child.[38]

5.35 The best interests of the child principle has been adopted in the proposed *Privacy Legislation Amendment (Enhancing Online Privacy and Other Measures) Bill 2021* (the Bill). The Bill's Explanatory Paper reflects that social media platforms pose a higher level of risk than other online services due to the popularity of social media services, the types of interactions that can happen, and the amount of personal information contained on social media services.[39] As a result, the Bill provides for the proposed Online Privacy code to require that social media and digital platforms:

> Ensure that the collection, use or disclosure of a child's personal information is fair and reasonable in the circumstances, with the best interests of the child

---

33 National Children's Commissioner, *Submission 64*, p. 5.

34 Ms Sarah Davies, AMF, *Committee Hansard,* 21 December 2021, p. 16.

35 ReachOut Australia, *Submission 36*, p. 11.

36 Orygen, *Submission 27.1*, p. 2.

37 Ms Sarah Davies, AMF, *Committee Hansard,* 21 December 2021, p. 16.

38 Ms Anne Hollands, National Children's Commissioner, AHRC, *Committee Hansard*, 2 March 2022, p. 1.

39 Attorney-General's Department (AGD), *Explanatory paper – Privacy Legislation Amendment (Enhancing Online Privacy and Other Measures) Bill 2021*, October 2021, available at: https://consultations.ag.gov.au/rights-and-protections/online-privacy-bill-exposure-draft/user_uploads/online-privacy-bill-explanatory-paper.pdf (accessed 9 February 2022), p. 11.

being the primary consideration when determining what is fair and reasonable.[40]

Implementation of National Principles for Child Safe Organisations

5.36 In February 2019, the Council of Australian Governments endorsed the National Principles for Child Safe Organisations (the Child Safe Principles), aimed at providing a 'nationally consistent approach to creating organisational cultures that foster child safety and wellbeing'.[41] These principles were recommended by the Royal Commission into Institutional Responses to Child Sexual Abuse, which recommended that all Australian institutions 'that engage in child-related work' be required to implement the standards.[42] The Child Safe Principles broadly cover not only child sexual abuse but other forms of harm to children and young people.

5.37 The Child Safe Principles consist of:

1 Child safety and wellbeing is embedded in organisational leadership, governance and culture;

2 Children and young people are informed about their rights, participate in decisions affecting them and are taken seriously;

3 Families and communities are informed and involved in promoting child safety and wellbeing;

4 Equity is upheld and diverse needs respected in policy and practice;

5 People working with children and young people are suitable and supported to reflect child safety and wellbeing values in practice;

6 Processes to respond to complaints and concerns are child focused;

7 Staff and volunteers are equipped with the knowledge, skills and awareness to keep children and young people safe through ongoing education and training;

---

40 AGD, *Explanatory paper – Privacy Legislation Amendment (Enhancing Online Privacy and Other Measures) Bill 2021*, October 2021, available at: https://consultations.ag.gov.au/rights-and-protections/online-privacy-bill-exposure-draft/user_uploads/online-privacy-bill-explanatory-paper.pdf (accessed 9 February 2022), p. 11.

41 AHRC, *About the National Principles*, available at: https://childsafe.humanrights.gov.au/national-principles/about-national-principles (accessed 28 January 2022).

42 AHRC, *About the National Principles*, available at: https://childsafe.humanrights.gov.au/national-principles/about-national-principles (accessed 28 January 2022).

8 Physical and online environments promote safety and wellbeing while minimising the opportunity for children and young people to be harmed;

9 Implementation of the national child safe principles is regularly reviewed and improved; and

10 Policies and procedures document how the organisation is safe for children and young people.[43]

5.38 Witnesses to the inquiry urged that digital platforms adopt the protections outlined in the National Principles for Child Safe Organisations. The National Children's Commissioner explained that Principle 8 was specifically directed at risks facing children in online environments, but that all of the principles were relevant for the digital industry.[44] The AMF similarly recommended that digital platforms be required to align with the National Principles for Child Safe Organisations.[45] This suggestion was also put by yourtown, who stated that the Child Safe Principles specifically refer to 'not just safe physical environments but safe online environments for children under 18 years of age'.[46]

5.39 yourtown posited that, given young people are spending increasing amounts of time online and that cyberbullying is increasing, online spaces which provide services to children should be required to adhere to child safety principles in a way that is understandable and accessible to its users.[47] yourtown also suggested that the incorporation of the National Principles into the Safety by Design principles would promote child safety in the design and development of online platforms from the outset.[48]

---

43 AHRC, *About the National Principles*, available at: https://childsafe.humanrights.gov.au/national-principles/about-national-principles (accessed 28 January 2022).

44 Ms Anne Hollands, National Children's Commissioner, AHRC, *Committee Hansard*, 2 March 2022, p. 1.

45 Ms Sarah Davies, AMF, *Committee Hansard,* 21 December 2021, p. 17.

46 Ms Kathryn Mandla, Head, Advocacy and Research, yourtown, *Committee Hansard*, 21 December 2021, p. 33.

47 Ms Kathryn Mandla, yourtown, *Committee Hansard,* 21 December 2021, p. 33.

48 Ms Kathryn Mandla, yourtown, *Committee Hansard,* 21 December 2021, p. 33.

Education and support

5.40 A key theme in evidence was the need for increased education for all citizens in relation to digital wellbeing, in particular for children and families. As discussed in Chapter 2 in the context of children and young people, however, the educational needs of different groups need to be targeted to avoid oversaturation in certain areas while neglecting others.

5.41 eSafety currently has a legislative mandate to 'coordinate education and online safety activities across the Commonwealth' and uses a 'holistic approach beyond prevention and protection'.[49] eSafety stated that it was of the view that digital education was a 'lifelong journey', particularly given that the age of digital use continues to get younger.[50]

5.42 eSafety provides a suite of educational resources on its website, aimed at different groups, such as:

- parents and carers as the front lines of defence, particularly in the early years
- educators and schools to develop students' critical skills across the 4 Rs (respect, resilience, responsibility and reasoning), to manage online safety incidents that may arise within the school community, and to support best practice online safety education
- domestic and family violence frontline workers to upskill people who support those experiencing technology-facilitated abuse, and
- specific diverse and vulnerable communities that our research shows are more likely to experience online harms.[51]

5.43 When reflecting on eSafety's educational services, witnesses were almost uniformly in support of the agency's educational materials, confirming that they were utilised successfully in varying contexts.[52]

5.44 Outside of government providers, multiple organisations provide educational programs to children and young people on digital literacy,

---

49 eSafety Commissioner, *Submission 53*, p. 3.

50 eSafety Commissioner, *Submission 53*, p. 7.

51 eSafety Commissioner, *Submission 53*, p. 8.

52 The Association of Heads of Independent Schools Australia, *Submission 24*, p. 6.

citizenship and safety. A range of providers who submitted to the inquiry included:

- The AMF;[53]
- Body Safety Australia;[54]
- The Carly Ryan Foundation (CRF);[55]
- The Daniel Morcombe Foundation; and
- Dolly's Dream.[56]

5.45 These groups all provide unique perspectives on online safety, which flows through in their education programs. For example, the AMF currently provides educational programs for young people on digital literacy, which focus on digital resilience:

> We deliberately take a strengths based approach because all of the evidence, not just here and not just recently but for decades around the world, around how you address disadvantage and vulnerability is about the importance of efficacy of building a strengths based approach. If you can help people identify and build their competencies, skills, strengths, resilience, EQ, IQ and now their digital quotient, DQ, it gives them the tools that mean they are much more likely to withstand the chipping away and undermining of their power and of their strengths to then succumb to what you've described around that coercion. That is deliberate, manipulative and very intentional. If you don't have, as a self, the recognition of those strengths that need then to be labelled as such and seen as such, it is much easier to chip away and undermine them. When we work in social change it is proven time and time again for individual wellbeing, resilience, strength and positive mental health that a strengths based approach is much more effective than talking to people about their deficits and problems.[57]

5.46 Education was seen as a critical component by many witnesses in reducing harm from online abuse. The Association of Heads of Independent Schools of Australia stated that education is a strong way of creating cultural change in addition to providing information and support to individuals.[58] The AMF too was supportive of a strong educational program focussing on digital

---

53 AMF, *Submission 2*.

54 Body Safety Australia, *Submission 59*.

55 The Carly Ryan Foundation (CRF), *Submission 54*.

56 Dolly's Dream, *Submission 4*.

57 Ms Sarah Davies, AMF, *Committee Hansard,* 21 December 2021, p. 17.

58 The Association of Heads of Independent Schools Australia, *Submission 24*, p. 5.

literacy, aimed at age groups from pre-school to the end of secondary school with developmentally appropriate content.[59] It also recommended digital literacy training for adults, particularly parents and carers, and for rural and regional communities.[60]

5.47 Among other suggestions for focus points of digital literacy education included:

- Disinformation and critical thinking skills when consuming online information;[61]
- Addressing the needs and challenges of parents with little or no digital literacy;[62]
- Educational materials for those with English as a second language or low proficiency in written English;[63] and
- Schools and teaching staff-specific training in dealing with situations involving social media or digital platforms.[64]

Developing appropriate and relevant educational programs

5.48 In expressing support for further education in relation to digital literacy, some providers cautioned that education in relation to online safety matters needed to be done carefully and consider the developmental stage of the age group that the education was directed. The AMF noted that responding to the issues relating to online safety required consideration of:

> the externalities or context within which children and young people are experiencing social media in a digital world, which is how essential it is to build strength in early childhood settings, school communities, families and those ancillary support services.[65]

5.49 Increased education for children and young people in relation to available redress and remediation was strongly recommended by yourtown. Ms Kathryn Mandla, Head of Advocacy and Research at yourtown, stated that 'having accessible ways to raise a concern, provide feedback, self-

---

59 Ms Sarah Davies, AMF, *Committee Hansard,* 21 December 2021, p. 21.

60 Ms Sarah Davies, AMF, *Committee Hansard,* 21 December 2021, p. 21.

61 CDW, *Submission 47*, p. 5.

62 Council of Catholic School Parents NSW ACT (CCSP), *Submission 32*, p. 3.

63 CCSP, *Submission 32*, p. 3.

64 Ms Sarah Davies, AMF, *Committee Hansard,* 21 December 2021, p. 20.

65 Ms Sarah Davies, AMF, *Committee Hansard,* 21 December 2021, p. 17.

advocate or make a complaint' would enhance transparency for young people and their advocates.[66]

5.50 Body Safety Australia also noted that education providers need to be conscious that a 'one size fits all' approach was not appropriate, and that programs needed to be carefully developed to meet the needs of the community:

> Education for children, young people and their families/carers is an essential component of harm minimisation. It requires expert knowledge, a trauma-informed perspective and a child-first framework. We warn against a one-size-fits-all approach to child protection. Culturally appropriate and age-appropriate education is essential to an effective education program, as is a whole-of-community approach. Children, young people, educator, families/carers, and other community groups all need to be included in preventative education approaches.[67]

5.51 Ms Davies of the AMF stated that education focussed on digital ethics and citizenship has been found most effective:

> I suppose one of our critical underlying philosophies is that consumers and people who use the digital world are not passive recipients. They are active agents in creating that space. So the more we can support children and young people to be positive digital citizens, coming back to that, understanding that everything they do and say not just curates their own experience but influences everyone else's and, build that sense that they're actually building other people's experience, then we are going to minimise and mitigate those experiences for other people. Recent research said that nine out of 10 Australian teens deliberately describe themselves choosing positive behaviours online. So the more we can get people to understand that it's not just their own experience that they're influencing and it's actually everyone else's, by building that awareness and that deeper understanding, I also think that then mitigates and minimises the negative.[68]

5.52 One potential model supported by some witnesses was the concept of a digital 'licence', directed at school-aged children, which focused on providing digital literacy skills prior to the award of the licence. Ms Kate Everett, Founder of Dolly's Dream, explained that it had worked with the AMF to create a digital licence in addition to a 'DigiPledge', an educational

---

66 Ms Kathryn Mandla, yourtown, *Committee Hansard,* 21 December 2021, p. 33.

67 Body Safety Australia, *Submission 59*, p. 4.

68 Ms Sarah Davies, AMF, *Committee Hansard,* 21 December 2021, p. 20.

program for families to complete together.[69] She explained the concept of the digital licence, stating:

> [W]e're not allowed to drive a car and we're not allowed to do so many other things in life without the appropriate education and licensing. It only seems logical that we would have something like this in the online world.[70]

Youth engagement in educational programs

5.53 Notwithstanding the need for effective educational strategies, evidence to the inquiry indicated that educational campaigns targeted at young people may not be as effective as possible and require improvement.

5.54 Research commissioned by the eSafety Commissioner suggested that young people believe that there is an 'oversaturation' of material at school and in the home in relation to cyberbullying materials in particular.[71] Similarly, research from the Young & Resilient Research Centre suggested that young people found that educational campaigns tended to be overly prohibitive and mis-informed, which resulted in their feeling 'misunderstood and disempowered by online safety conversations'.[72]

5.55 The research further indicated that young people were most concerned about particular issues in relation to online safety. The top three concerns reported by young people were identified as:

- Interactions with people online (such as catfishing, fake accounts and contact from unknown people);
- Privacy matters (including exposure of personal data, photos and stolen identities); and

---

69 Ms Kate Everett, Founder, Dolly's Dream, *Committee Hansard,* 27 January 2022, p. 5.

70 Ms Kate Everett, Founder, Dolly's Dream, *Committee Hansard,* 27 January 2022, p. 5.

71 Young & Resilient Research Centre, *Consultations with young people to inform the eSafety Commissioner's Engagement Strategy for Young People*, 2021, available at: https://www.esafety.gov.au/sites/default/files/2022-01/YRRC%20Research%20Report%20eSafety%202021_web%20V06%20-%20publishing_1.pdf (accessed 28 January 2022), p. 6.

72 Young & Resilient Research Centre, *Consultations with young people to inform the eSafety Commissioner's Engagement Strategy for Young People*, 2021, available at: https://www.esafety.gov.au/sites/default/files/2022-01/YRRC%20Research%20Report%20eSafety%202021_web%20V06%20-%20publishing_1.pdf (accessed 28 January 2022), p. 5.

- Security issues (such as hacking, scams and malware).[73]

5.56 The findings identified six best-practice principles for youth engagement, as suggested by the participants of the study:

- Incorporating diversity and inclusivity in engagement programs;
- Ensuring programs are youth-led while being supported by adults where required;
- Genuinely listening to young people and formulating active solutions or outcomes;
- Collaboration with young people in designing, delivering and evaluating programs and policies targeted at their age group;
- Providing benefits to engagement to promote learning, growth and development, in addition to fair compensation; and
- Integrating fun, age-appropriate and accessible activities for youth engagement.[74]

5.57 The CRF provided anecdotal reports that its consultations with young people had identified that current educational methods do not adequately understand or address young people's concerns regarding online harm.[75] For example, the CRF stated that it had found that young people often used online pornography to learn about relationships and sexual activity, often because the sexuality education provided in schools did not reflect their lived experiences or concerns. This had the potential to result in damaging understandings about sexuality and relationships, particularly in relation to violent or aggressive behaviour.[76]

5.58 The National Mental Health Commission (NMHC) similarly suggested that it found that young people often can recognise danger or where they feel unsafe in online platforms, but that they were not being supported fully to

---

73 Young & Resilient Research Centre, *Consultations with young people to inform the eSafety Commissioner's Engagement Strategy for Young People*, 2021`, available at: https://www.esafety.gov.au/sites/default/files/2022-01/YRRC%20Research%20Report%20eSafety%202021_web%20V06%20-%20publishing_1.pdf (accessed 28 January 2022), p. 12.

74 Young & Resilient Research Centre, *Consultations with young people to inform the eSafety Commissioner's Engagement Strategy for Young People*, 2021, available at: https://www.esafety.gov.au/sites/default/files/2022-01/YRRC%20Research%20Report%20eSafety%202021_web%20V06%20-%20publishing_1.pdf (accessed 28 January 2022), p. 6.

75 CRF, *Submission 54*, p. 8.

76 CRF, *Submission 54*, p. 8.

manage the risks. The NMHC stated that, in looking at ways in which to best support young people online, it takes a broad view of young people and how they use the internet:

> This isn't just about how we are managing content online or how we are providing them with the equivalent of the old-fashioned media literacy. I think we need to go much deeper than that. We need to be asking: what patterns of behaviour are our young people exhibiting? Are they any different to how they've been before? What are the consequences on their behaviours of being in an online environment which has two attributes that may not have been there in generations past? One is a plethora of information in real time and much deeper, broader information than was available before. The second is an urgency and an immediacy about decision-making that has not been there in past generations. They're the types of issues that I think we need to look at through a professional discipline lens and ask what we could or should be doing in that space.[77]

5.59 eSafety stated that, in developing educational materials for young people, it had a multifaceted approach targeting the specific needs and wants of young people in terms of digital literacy; eSafety has a youth engagement strategy, in addition to the recent establishment of its Online Safety Youth Advisory Council.[78] eSafety reported that their efforts to engage with children and young people appear to be making a difference in how they addressed online safety:

> Over the last several years, we have started to see evidence of real change to behaviours and attitudes, with children and young people taking multiple actions and accessing a range of tools and tactics in response to negative experiences. For example, eSafety's 2021 survey of 3,600 young Australians aged 8-17 years-old found that 64 per cent of young people who have experienced negative online behaviour blocked or unfriended people who had bullied them online – a significant increase on 46 per cent of young people in 2017. The research also found that young people are increasingly reaching out to their parents and friends. Sixty-six per cent of young people who have experienced negative online behaviour told their parents (up from 55% in 2017) and 60% told their friends (up from 28% in 2017).[79]

---

77 Ms Christine Morgan, Chief Executive Officer and Prime Minister's National Suicide Prevention Adviser, National Mental Health Commission (NMHC), *Committee Hansard,* 21 January 2022, p. 9.

78 eSafety Commissioner, *Submission 53*, p. 8.

79 eSafety Commissioner, *Submission 53*, pp 8-9.

5.60 The National Children's Commissioner advocated that children and young people have the right to be consulted and their views to be appropriately considered in any reform process which impacts their lives.[80] She explained that the UN has recognised the importance of not allowing adults to assume what children's needs are, and that an overly restrictive approach can be developmentally stunting in terms of autonomy and independence.[81]

Parental and community education of online risks

5.61 Witnesses advocated for greater emphasis to be placed on providing parents and carers with sufficient educational sources to effectively manage their children's usage and respond appropriately where issues arise.

5.62 A number of witnesses raised concerns that parents did not sufficiently understand the dangers or risks posed to online users. Dr Jessie Mitchell, Advocacy Manager at the AMF, pointed to recent research conducted by the Australian Centre to Counter Child Exploitation which identified that, for parents and carers, there are:

> … still found quite low levels of understanding of how child sexual exploitation and abuse operates online, the severity of it and how easy that sometimes is to occur. There was a sense that, while parents were relatively across what they needed to do to keep children safe face to face, there was a certain lack of understanding of how those behaviours could function in a digital world.[82]

5.63 This point was similarly raised by the Council of Catholic School Parents NSW ACT (CCSP), which stated that parents do not necessarily have the appropriate understanding of the risks of online engagement for their children. The CCSP stated:

> Not all parents are digitally literate or have a full understanding of the potential short- and long-term risks that children may face from using social media platforms. This extends to their digital identity/footprint, personal reputation, viewing of age-inappropriate content and unknowingly breaking the law by possessing and forwarding certain images.[83]

---

80 National Children's Commissioner, *Submission 64*, p. 5.

81 National Children's Commissioner, *Submission 64*, p. 5.

82 Dr Jessie Mitchell, Advocacy Manager, AMF, *Committee Hansard,* 21 December 2021, p. 18.

83 CCSP, *Submission 32*, p. 3.

5.64 The CCSP further provided anecdotal detail of parents unwittingly assisting children to engage in risky behaviour, including setting up a social media account for underage children, and 'turn a blind eye' to how children are engaging with online platforms.[84] In addition, the CCSP noted that some parents are unaware of how their own behaviour on their social media accounts may impact their children, such as publishing photos of their children on their personal accounts.[85]

5.65 Some witnesses argued that providing education to adults would assist in dealing with online abuse for children and young people. The NMHC argued that, for young people in particular, engaging with their 'circles of influence' – parents, carers, family members, and others who impact them – and providing adequate resources and support for these particular groups, would assist in reducing harm.[86]

5.66 Concerningly, witnesses told the Committee that large sections of the community were not aware of the powers of the eSafety Commissioner and other government agencies in addressing online harm. The AMF stated that there were low levels of understanding in the community about reporting mechanisms in relation to online abuse, expressing the view that the work of eSafety and other regulatory bodies was not widely known or understood. Dr Mitchell stated:

> I think part of the difficulty may be at the other end: the public's knowledge around the fact that there are places where they can report concerns, whether that's to the eSafety Commissioner or to the ACCCE or to Scamwatch or somewhere like that. There's still, I think, not sufficiently high and even understanding of those reporting options in the community, particularly amongst families where there might be lower levels of digital literacy and particularly if there's a family where perhaps the parents or carers are not very confident online and don't necessarily have the digital literacy to respond quickly when something goes wrong. So I think there is work to be done to make those avenues for reporting more accessible and recognised by all Australians.[87]

5.67 This statement was echoed by the eSafety Commissioner, who explained at a Senate Additional Estimates hearing on 15 February 2022 that many social

---

84 CCSP, *Submission 32*, p. 3.

85 CCSP, *Submission 32*, p. 3.

86 Ms Christine Morgan, NMHC, *Committee Hansard,* 21 January 2022, p. 6.

87 Dr Jessie Mitchell, AMF, *Committee Hansard,* 21 December 2021, p. 18.

media companies provide training to police forces in relation to criminal compliance provisions in relation to their platforms, but that this was not necessarily reflected at the local level of law enforcement.[88] The eSafety Commissioner further noted that eSafety works with law enforcement agencies at state and territory levels, and were working to establish agreements and education campaigns with police commissioners to improve awareness of eSafety's powers.[89]

5.68 In relation to the levels of understanding of local law enforcement entities, Body Safety Australia recommended educational programs be expanded to include entities such as police and courts to address in training law enforcement officials in identifying and appropriately managing online harm matters.[90]

Education is not the 'silver bullet' in addressing online harm

5.69 The enthusiasm demonstrated by witnesses for improved education was tempered by warnings not to consider education as the 'silver bullet' in addressing online harm. Education is only part of the answer to increasing online safety. Dr Salter raised concerns that while education (and in particular parental education) is important, it should not be considered as the primary mode of defence against online harms, nor perpetuate the idea that young people and families are responsible for online safety:

> Of course parental education will always be part of the online safety equation. There's no question that parents of course have a role in the safety of their children. It's also quite sensible to provide education to children about services and so on. The issue is that, over the last 25 years, that's been our primary bulwark to keep children safe. We know that it's not working, and we know that it cannot work. The example I give you is if we had no child safety standards for child playgrounds and if child playgrounds were filled with sharp edges, bits of metal and plastic that were unstable and so on. It wouldn't matter how much education we gave children and adults about how to play safe in playgrounds. If the playgrounds were not safe, children would come to harm in those playgrounds. That is the case in the online environment. These

---

88 Ms Julie Inman Grant, eSafety Commissioner, Office of the eSafety Commissioner, Additional Estimates 2021-22, Senate Environment and Communications Legislation Committee, *Committee Hansard*, 15 February 2022, p. 25.

89 Ms Julie Inman Grant, eSafety Commissioner, Office of the eSafety Commissioner, Additional Estimates 2021-22, Senate Environment and Communications Legislation Committee, *Committee Hansard*, 15 February 2022, p. 29.

90 Body Safety Australia, *Submission 59*, p. 4.

> environments are fundamentally unsafe at the moment. So, unfortunately, the focus on parental education can act as a sort of a distraction from the structural issues that ensure that children are unsafe now and into the future.[91]

5.70 Dr Salter further put the view that education does not overcome the key problem that 'these platforms are fundamentally unsafe'.[92] He noted that relying on education did not reflect the reality that many parents are either unable for various reasons to monitor their children's online usage, or are the perpetrators of online child exploitation themselves.[93] He stated:

> The other point I want to make is that when we look at child sex abuse material we see that a lot of that content is in fact produced in the home by sexually abusive fathers and family members. So what do we do for those kids who are unsafe because of their parents? It's their parent that is exploiting them. When we look on the dark web and online we see a lot of men who are abusers and are talking about the fact that they're abusing their child. As a community and as a country we have to get real about the range of risks that are posed to children. Absolutely we want to empower parents, but, unfortunately, they can't be the front line of defence. Child protection is a collective responsibility. That responsibility should be shared by online service providers and of course articulated and enforced by government.[94]

Government leadership in addressing online safety

5.71 Witnesses emphasised the need for government at all levels to be a leader in the promotion of online safety measures. yourtown stated that it viewed government as having a role in multiple ways:

> There is the regulatory role. There is a leadership role in setting the culture and the standard and role modelling best-practice behaviour and creating public awareness of the harms that can come about through poor safety design for children and young people and having poor processes and policies in place that can harm children and young people.[95]

5.72 yourtown also suggested that further collaboration and leadership between the Commonwealth and the state and territory governments would improve

---

91 Dr Michael Salter, *Committee Hansard,* 18 January 2022, p. 14.

92 Dr Michael Salter, *Committee Hansard,* 18 January 2022, p. 14.

93 Dr Michael Salter, *Committee Hansard,* 18 January 2022, pp 14-15.

94 Dr Michael Salter, *Committee Hansard,* 18 January 2022, p. 14.

95 Ms Kathryn Mandla, yourtown, *Committee Hansard,* 21 December 2021, p. 33.

the effectiveness of programs. Dr Marion Byrne of yourtown stated that issues in relation to the implementation of agreed principles at local levels and disconnected systems could potentially cause dysfunction when attempting to reduce harm.[96]

Committee comment

5.73 Social media platforms occupy an unprecedented position in modern society. As Mr Matt Berriman stated, social media platforms are no longer just a form of technology but are an 'infrastructure business' where they are no longer optional but a necessity for many users to conduct various aspects of their lives and – if turned off – can be extraordinarily detrimental to people's personal and professional lives.[97]

5.74 One of the key challenges for all participants in this sector is recognising the importance of balancing the freedom of online participants to engage in democratic debate and express their views, and the need for protection for the most vulnerable users of online spaces. The Committee is conscious that its observations and recommendations may have significant ramifications for this balance.

5.75 Having said that, it is clear that the status quo cannot continue. The Committee supports the view expressed by Ms Frances Haugen and others that social media platforms, in addition to other digital services, have not demonstrated their willingness to put the safety of their users before other considerations. While privacy concerns are critical to the rights of all internet users, the Committee does not believe that these issues outweigh the fundamental issue of ensuring safety in online environments.

5.76 Indeed, during the course of this inquiry, Meta released the company's new values, one of which, 'Move Fast', which includes the following:

> Move Fast helps us to build and learn faster than anyone else. This means acting with urgency and not waiting until next week to do something you could do today. At our scale, this also means continuously working to increase

96 Dr Marion Byrne, Manager, Advocacy, Research and Innovation, yourtown, *Committee Hansard*, 21 December 2021, p. 33.

97 Mr Matt Berriman, Chair, Mental Health Australia, *Committee Hansard*, 21 January 2022, p. 22.

> the velocity of our highest priority initiatives by methodically removing barriers that get in the way.[98]

5.77 Despite the assurances of the social media giants that have provided evidence to this inquiry that they take safety concerns seriously, the Committee is unable to reconcile these statements with the demonstrated attitude of Meta and others in the social media industry which appear to value progress, pace and profit above all other concerns.

5.78 The time has come to fundamentally shift the burden of responsibility regarding ensuring online safety. For too long, the onus of maintaining online safety has been on the most vulnerable users, including children and their parents. This is unacceptable and unsustainable in an environment where users like children are exposed to the most risk online and suffer extreme forms of harm as a result.

Improving online discourse

5.79 When considering standards of acceptable behaviour and commentary online, the need to ensure freedom of speech must be weighed against a user's right to be safe. Social media platforms ultimately make a judgement as to how these objectives are effectively balanced.

5.80 It is the Committee's view that these companies do not always achieve the most appropriate outcome. Given the prevalence of online communication in the lives of Australians, this in turn impacts community standards of what conduct is deemed acceptable in society more broadly.

5.81 This can be better addressed if social media platforms place a stronger focus on a user's right to be safe online, and clearly indicate that respectful dissent and disagreement will be tolerated, while abuse will not.

A statutory duty of care framework

5.82 A statutory duty of care was suggested by a number of witnesses and has previously been recommended by other parliamentary committees. The Committee notes the approach taken by the UK Government, which has been to 'call time' on social media companies' system of co- or self-regulation, evident in the draft legislation governing its new approach to online safety matters. The Committee particularly notes the argument that the use of a statutory duty of care would ensure that social media platforms

---

98 Meta, *Culture at Meta*, available at: https://www.metacareers.com/facebook-life/ (accessed 28 February 2022).

are responsible for creating systems which are protective and responsive to their users. This model has significant strengths, and flips the onus of responsibility to provide and ensure user safety back onto social media platforms.

5.83 The Committee supports the proposed model of a formalised framework for duty of care and believes it would be an enhancement to the existing Australian regulatory framework of the Basic Online Safety Expectations. A framework for a duty of care would assist in ensuring that digital platforms have an incentive to create systems, and improve current ones, to ensure the safety of all users, particularly children, women, and other vulnerable groups. A formalised duty of care would also ensure that such a model is incorporates penalties for non-compliance.

5.84 An introduction of a such a framework would require further amendments to the Basic Online Safety Expectations to ensure its strength as key requirement for industry actors. However, the Committee is conscious of the OSA's short time in operation, and other ongoing legislative reviews. Further, the eSafety Commissioner's Safety by Design program may go some way in addressing the concerns raised in relation to the need for compliance with a formalised framework for duty of care.

5.85 It would therefore be appropriate to consider the need for the implementation of such enhancements to the duty of care framework within the Basic Online Safety Expectations for social media and other digital platforms in a broad-scale review of the entire digital industry and its overarching legislative framework. The review could consider how the principles of duty of care would address any potential gaps left unregulated by the OSA, and also consider how a the Basic Online Safety Expectations framework could incorporate the duty of care within its model.

5.86 Further, a formal duty of care framework should incorporate the best interests of the child principle as its guiding model, which will ensure online platforms place this concept at the forefront when designing new products and updating existing services.

5.87 The Committee also supports the implementation of the National Principles for Child Safe Organisations for digital media platforms; however, it observed in evidence that these principles are designed to apply to digital platforms already. A more clearly articulated application of these principles should be considered in any further reform in relation to the protection of children online.

Recommendation 20

5.88 The Committee recommends that the Digital Review include in its terms of reference:

- **The need to strengthen the Basic Online Safety Expectations to incorporate and formalise a statutory duty of care towards users;**
- **The scope and nature of such a duty of care framework, including potential models of implementation and operation;**
- **Potential methods of enforcement to ensure compliance, including penalties for non-compliance; and**
- **The incorporation of the best interests of the child principle as an enforceable obligation on social media and other digital platforms, including potential reporting mechanisms.**

Education

5.89 The Committee recognises the urgent need for education for large parts of the community, particularly for the most vulnerable users. Having said that, the eSafety Commissioner, in addition to the non-government sector, currently provides extensive forms of education materials. It is not enough, therefore, to broadly state that more education is needed. To ensure that educational campaigns are effective, they must be targeted and carefully designed to avoid oversaturating recipients or focusing on material that is irrelevant to people's lived experiences.

5.90 Young people in particular feel that they are oversaturated with material aimed at their age group, while they are also often the most at-risk and experience high proportions of harm. This may indicate that the current educational material is not sufficient in addressing the specific needs of young people in regards to online safety.

5.91 The eSafety Commissioner's work in attempting to better understand the needs of young people and how best to engage with them is therefore encouraging. The implementation of these findings in eSafety's educational campaigns will, in the Committee's view, be indicative of where further work is needed.

Recommendation 21

5.92 The Committee recommends that the eSafety Commissioner:

- **Increase the reach of educational programs geared at young people regarding online harms, with a particular focus on reporting mechanisms and the nature of some online harms being a criminal offence;**

- **Formalise a consultation and engagement model with young people through the Australian Government's Youth Advisory Council in regards to educational themes and program delivery; and**

- **Report to the Parliament on the operation and outcomes of the program, including research identifying whether this has resulted in a reduction in online harm for young people.**

5.93 The Committee also agrees with the concerns raised by many witnesses to the inquiry suggesting that education more broadly does not sufficiently address digital literacy and safety. It also noted the observations of witnesses that education needed to begin from early childhood and beyond secondary schooling years. Separate to these concerns, however, it is clear that a large majority of the adult population similarly do not have the digital literacy skills needed to manage online interaction safely.

5.94 It a necessary and elemental step of Australia's fight against online harm to invest in and provide digital education for all Australians, in all community groups, particularly as the digital world develops. Educational programs should be expanded to a much wider range of groups, including the most vulnerable users in society.

5.95 Given that education is required at all age groups and areas of society, the methods in which education can be provided should be considered carefully and with understanding of intersectional factors such as developmental stages, cultural differences, language barriers and access to technology. The eSafety Commissioner should work with relevant government agencies and departments to identify the most appropriate form of implementing educational campaigns to society, including inclusion into the National Curriculum.

5.96 While the Committee heard evidence in relation to the prospect of the widescale implementation of a digital licence for children and young people,

it does not support the mandatory use of a digital licence for children to access the internet. It does, however, believe that a digital licence could be used as an educational tool aimed at school-aged children, similarly to the way that children are issued 'pen licences'.

5.97 The Committee also observed that many people in the community were not aware of the powers of the eSafety Commissioner in removing harmful content. This should be rectified as a matter of urgency, as it is arguably of little use for the eSafety Commissioner to have these powers if people are unaware of them.

5.98 The Committee further suggests that the Australian Government continue to provide a leadership role in working with states and territories to implement digital education programs at local levels.

Recommendation 22

5.99 The Committee recommends that the eSafety Commissioner work in consultation with the Department of Education, Skills and Employment to design and implement a national strategy on online safety education designed for early childhood, and primary school-aged children, and secondary school-aged young people, including:

- **A proposed curriculum, informed by developmental stages and other relevant factors;**
- **Potential methods of rollout, including consultation and engagement with children, young people, child development and psychology experts, digital education experts and other specialists in online harm; and**
- **A roadmap provided to parents of these age groups detailing methods of addressing online harm.**

Recommendation 23

5.100 The Committee recommends that the eSafety Commissioner design and administer an education and awareness campaign aimed at adults, particularly in relation to vulnerable groups such as women, migrant and refugee groups, and people with disabilities, with a focus on the eSafety Commissioner's powers to remove harmful content and the mechanisms through which people can report harmful content and online abuse.

Recommendation 24

5.101 The Committee recommends that the Australian Government work with states and territories to ensure that relevant law enforcement agencies are appropriately trained on how to support victims of online harm. This should include trauma-informed approaches as well as a comprehensive understanding of police powers and other relevant avenues, such as the relevant powers of the eSafety Commissioner.

Recommendation 25

5.102 The Committee recommends that the Australian Government review funding to the eSafety Commissioner within twelve (12) months to ensure that any of the Committee's recommendations that are agreed to by the Government and implemented by the Office of the eSafety Commissioner are adequately and appropriately funded for any increased resource requirements.

Engagement with young people

Recommendation 26

5.103 The Committee recommends that the Online Safety Youth Advisory Council, via the eSafety Commissioner, provide a response to this report and its recommendations within six (6) months of its establishment and full membership.

Lucy Wicks MP
Chair

11 March 2022

A. Submissions

1 Adam Johnston

2 Alannah & Madeline Foundation

3 Australian Muslim Advocacy Network

4 Dolly's Dream

5 Cyber Security CRC

6 The Australia Institute - Centre for Responsible Technology

7 Dr Emily van der Nagel

8 Tasmanian Catholic Schools Parents Council

9 Australian Small Business and Family Enterprise Ombudsman

10 Butterfly Foundation

11 Australian Communications and Media Authority

12 Reset Australia

13 Cr Jack Dempsey

14 Australian Communications Consumer Action Network

15 Family Zone

16 Snap Inc.

- 16.1 Supplementary to submission 16

17 Dr Bridget Harris, members of The Independent Collective of Survivors, Molly Dragiewicz and Delanie Woodlock

18 Mr Gerard Hosier

19 Victorian Health Promotion Foundation (VicHealth)

20 Carnegie UK

21 Ms Nell McGill

22 Mr Benjamin Cronshaw

23 Digital Rights Watch

24 Association of Heads of Independent Schools of Australia

25 WESNET

26 Office of the Children's Commissioner Northern Territory

27 Orygen

28 Ms Nicole Shackleton

29 Media Diversity Australia

30 Google Australia

- 30.1 Supplementary to submission 30
- 30.2 Supplementary to submission 30

31 yourtown

32 Council of Catholic School Parents NSW/ACT

33 SBS

34 Harmony Alliance

35 Centre for Excellence in Child and Family Welfare

36 ReachOut Australia

- 36.1 Supplementary to submission 36

37 Eating Disorders Families Australia

38 Save the Children and Child Wise

39 Australian Community Managers

40 Department of Home Affairs

- 40.1 Supplementary to submission 40

41 Royal Australian College of General Practitioners (RACGP)

42 Free TV Australia

43 Giggle

44 Department of Infrastructure, Transport, Regional Development and Communications

- 44.1 Supplementary to submission 44

45 QUT Digital Media Research Centre

46 Digital Industry Group Inc (DIGI)

- 46.1 Supplementary to submission 46

47 Centre for Digital Wellbeing

48 Michael Douglas

49 Meta

- 49.1 Supplementary to submission 49
- 49.2 Supplementary to submission 49

50 Twitter

- 50.1 Supplementary to submission 50

51 Dr Cassandra Cross

52 The Synod of Victoria and Tasmania, Uniting Church in Australia

53 eSafety Commissioner

- 53.1 Supplementary to submission 53
- 53.2 Supplementary to submission 53

54 The Carly Ryan Foundation

55 Australian Research Alliance for Children & Youth

56 Australian Council on Children and the Media

57 TikTok Australia

- 57.1 Supplementary to submission 57

58 National Mental Health Commission

- 58.1 Supplementary to submission 58

59 Body Safety Australia

60 *Name Withheld*

61 *Name Withheld*

62 Nyadol Nyuon

- 62.1 Supplementary to submission 62

63 *Confidential*

64 National Children's Commissioner

65 Dr Michael Salter

66 Professor David Flint

67 Isolated Children's Parents' Association of Australia

68 Collective Shout

- 68.1 Supplementary to submission 68

69 Senator Claire Chandler

70 Nicolle Flint MP

71 Attorney-General's Department

- 71.1 Supplementary to submission 71

72 *Confidential*

73 *Confidential*

74 Croakey Health Media

75 *Name Withheld*

76 Mr Paul Stewart

77 Law Council of Australia

78 Independent Schools Australia

79 Public Health Association of Australia

80 *Name Withheld*

81 Obesity Policy Coalition

82 Mr Nicholas Butler

83 *Name Withheld*

84 Miss Tracey Hoolachan

85 Scarlet Alliance, Australian Sex Workers Association

86 Dr Marshall Ballantine-Jones

87 Gender Equity Victoria

88 Fighters Against Child Abuse Australia

89 Feminist Legal Clinic Inc

90 *Confidential*

B. Exhibits

Centre for Excellence in Child and Family Welfare

Submission to the Senate Inquiry into the harm being done to Australian children through access to pornography on the internet, 16 February 2016

ReachOut Australia

Children and Young People's Mental Health and Wellbeing coalition's submission in response to the Privacy Legislation Amendment (Enhancing Online Privacy and Other Measures) Bill 2021, November 2021

Ginger Gorman

Manuscript of book, Troll Hunting

Michael Douglas

Submission on the Exposure Draft of the Social Media (Anti-Trolling) Bill, 21 January 2022

Katharine Gelber and Rita Jabri Markwell

Submission to the Parliamentary Joint Committee on Intelligence and Security, 7 May 2021

Dr Michael Salter

Open letter to Mark Zuckerberg, Facebook's approach to child protection and well-being

eSafety Commissioner

eSafety submission to the Social Media (Anti-Trolling) Bill

eSafety Commissioner

eSafety submission to the Consultation for Online Privacy Bill

Professor David Flint

Annexure: Comment on Craig Kelly MP And Privilege (Updated)

Baroness Kidron

5 Rights Foundation: The Age Appropriate Design Code

Baroness Kidron

5 Rights Foundation: IEEE 2089-2021 Standard for an Age Appropriate Digital Services Framework, January 2022

Office of the Australian Information Commissioner

OAIC's submission to the Attorney-General's Department consultation on the exposure draft of the Social Media (Anti-Trolling) Bill 2021

C. Public Hearings

Tuesday, 21 December 2021 - Canberra

The Carly Ryan Foundation

- Ms Sonya Ryan

Daniel Morcombe Foundation

- Mr Bruce Morcombe, Founder
- Ms Tracey McAsey, Manager
- Mrs Rebecca Borradale, Multimedia Designer

Alannah & Madeline Foundation

- Ms Sarah Davies, CEO
- Ms Ariana Kurzeme, Director, Policy & Prevention
- Dr Jessie Mitchell, Advocacy Manager

Young & Resilient Research Centre, Western Sydney University

- Dr Amanda Third, Professorial Research Fellow, Institute for Culture and Society; Co-Director, Young and Resilient Research Centre

yourtown (Kids Helpline)

- Ms Kathryn Mandla, Head of Advocacy and Research
- Ms Marion Byrne, Advocacy, Research and Innovation Manager

Wednesday, 22 December 2021 - Canberra

Carly Findlay OAM, Private capacity

Nyadol Nyuon, Private capacity

Executive Council of Australian Jewry

- Dr Andre Oboler, CEO and Managing Director, Online Hate Prevention Institute
- Mr Peter Wertheim, co-CEO

Australian Muslim Advocacy Network

- Ms Rita Jabri-Markwell, Advisor

Let Her Speak campaign

- Ms Nina Funnell, Founder
- Mr Michael Bradley, Campaign Partner

Tuesday, 18 January 2022 - Canberra

Erin Molan, Private capacity

Dr Michael Salter, Private capacity

Ms Noelle Martin, Private capacity

Thursday, 20 January 2022 - Canberra

Google Australia

- Ms Lucinda Longcroft, Director, Government Affairs and Public Policy, Google Australia and New Zealand
- Ms Samantha Yorke, Government Affairs and Public Policy, Google Australia

Meta (Facebook/Instagram)

- Mr Josh Machin, Head of Public Policy, Australia
- Ms Mia Garlick, Regional Director for Policy, Australia, New Zealand and the Pacific Islands

TikTok

- Mrs Julie de Bailliencourt, Global Head of Product Policy
- Mr Lee Hunter, General Manager, Australia and New Zealand
- Mr Brent Thomas, Director of Public Policy, Australia and New Zealand

Digital Industry Group Inc (DIGI)

- Ms Sunita Bose, Managing Director
- Dr Jennifer Duxbury, Director, Policy, Regulatory Affairs and Research

Professor Edward Santow, Industry Professor – Responsible Technology, University of Technology Sydney

Mr Michael Douglas, Private Capacity

Reset Australia

- Dr Rys Farthing, Director of Data Policy
- Ms Dhakshayini, Director of Tech Policy

Australian Communications and Media Authority

- Ms Nerida O'Loughlin, Chair and Agency Head
- Ms Cathy Rainsford, General Manager, Content and Consumer Division

Friday, 21 January 2022 - Canberra

National Mental Health Commission

- Ms Christine Morgan, Chief Executive Officer and Prime Minister's National Suicide Prevention Adviser
- Ms Lyndall, Deputy Chief Executive Officer

Twitter

- Ms Kara Hinesley, Director of Public Policy, Australia and New Zealand
- Ms Kathleen Reen, Senior Director of Public Policy, Asia-Pacific

Mental Health Australia

- Dr Leanne Beagley, Chief Executive Officer
- Mr Matt Berriman, Chair

Centre for Digital Wellbeing

- Ms Carla Wilshire, Chair
- Ms Veronica Finn, Manager

Thursday, 27 January 2022 - Canberra

Dolly's Dream

- Ms Kate Everett, Founder
- Mr Stephen Bendle, Manager

Centre for Excellence in Child and Family Welfare

- Ms Deborah Tsorbaris, Chief Executive Officer
- Dr Michele Lonsdale, Deputy Chief Executive Officer, Director Social Policy & Research

Isolated Children's Parents' Association of Australia

- Ms Alana Moller, President

Catholic School Parents Australia

- Mr John O'Brien, Executive Officer
- Ms Siobhan Allen, CSPA Executive Member

Family Voice

- Mr Peter Stevens, Victoria Director

Collective Shout

- Ms Melinda Tankard Reist, Movement Director
- Lyn Swanson Kennedy, Campaigner

Butterfly Foundation

- Ms Alex Cowen, Communications Manager
- Dr Sarah Squire, Manager – Knowledge, Research and Policy

Eating Disorders Families Australia

- Mr David Quilty, Director

WeProtect Global Alliance

- Mr Iain Drennan, Executive Director

Damien Collins MP, Chair of Joint Committee on the Draft Online Safety Bill, UK Parliament

Professor Lorna Woods, Professor of Internet Law, University Essex

Mr William Perrin, Trustee, Carnegie UK

Baroness Beeban Kidson, Member of House of Lords, UK Parliament

Friday, 28 January 2022 - Canberra

Ms Karen Bentley, Chief Executive Officer, WESNET

Ms Padma Raman, Chief Executive Officer, Australia's National Research Organisation for Women's Safety (ANROWS)

Dr Hany Farid, Private Capacity

Miss Sall Grover, Founder & Chief Executive Officer, Giggle

Centre for Responsible Technology

- Mr Peter Lewis, Director

Office of the Australian Information Commissioner

- Ms Angelene Falk, Information Commissioner and Privacy Commissioner Australia
- Ms Rebecca Brown, Acting Director, Law Reform & Government

Department of Infrastructure, Transport, Regional Development and Communications

- Mr Richard Windeyer, Deputy Secretary
- Ms Pauline Sullivan, First Assistant Secretary, Online Safety, Media and Platforms
- Ms Bridget Gannon, Assistant Secretary, Online Safety

Attorney General's Department

- Ms Julia Galluccio, Assistant Secretary - Information Law Branch, Integrity and Security Division
- Mr Michael Johnson, Assistant Secretary, Defamation Taskforce
- Mr Graham Bannerman, Director, Defamation Taskforce

Mr Brandon Silverman, Private capacity, Former Head of CrowdTangle

Tuesday, 1 February 2022 - Canberra

Department of Home Affairs

- Mr Brendan Dowling, First Assistant Secretary, Digital and Technology Policy Division
- Mr Peter Anstee, Assistant Secretary, Technology Policy Branch
- Ms Ciara Spencer, First Assistant Secretary, Law Enforcement Policy

Australian Football League – Executive

- Ms Kate Hall, Head of Mental Health and Wellbeing
- Ms Tanya Hosch, Executive General Manager Inclusion and Social Policy

Australian Football League – Players

- Ms Tayla Harris
- Mr Chad Wingard

Thursday, 3 February 2022 - Canberra

Ms Frances Haugen, Private capacity

eSafety Commissioner

- Ms Julie Inman Grant, Commissioner
- Mr Toby Dagg, Executive Manager, Investigations
- Ms Kelly Tallon, Manager, International, Strategy & Futures

Wednesday, 2 March 2022 – Canberra

Australian Human Rights Commission

- Ms Anne Hollonds, National Children's Commissioner

Snap Inc

- Mr Henry Turnbull, Head of Public Policy, Asia Pacific

Family Zone Cyber Safety Ltd

- Mr Timothy Levy, Managing Director

Meta (Facebook/Instagram)

- Mr Josh Manchin, Head of Public Policy
- Ms Mia Garlick, Director of Policy

Tuesday, 8 March 2022 – Canberra

Australian Centre to Counter Child Exploitation

- Commander Hilda Sirec, Australian Centre to Counter Child Exploitation, and Human Exploitation, Australian Federal Police
- Assistant Commissioner Lesa Gale, Northern Command, Australian Federal Police

Tuesday, 8 March 2022 – Virtual site visit

Virtual site visit with students from The Gap State High School

Wednesday, 9 March 2022 – Virtual site visit

Virtual site visit of TikTok's Transparency and Accountability Centre

Labor members' additional comments

While time constraints of this Inquiry did not permit a line-by-line negotiation of the final report between committee members, Labor members of the Committee support the intent of the final Committee Report and its recommendations.

Labor members thank government members, particularly the Chair, for the bipartisan spirit in which this inquiry was undertaken and the efforts that were made to ensure that issues raised by Labor members were heard in this committee process.

The harms currently being caused by social media platforms are by now well known. The era of effective self-regulation by these platforms failed to deliver a safe online environment for Australians. All sides of politics are united in the need for government regulation to underpin the safety of these platforms, particularly for vulnerable groups.

Labor has a strong record of supporting online safety measures for Australians, both in government and from opposition, and Labor members have sought to contribute to this inquiry on a constructive, bipartisan basis.

While supporting the Committee Report's recommendations, Labor members wish to make the following Additional Comments.

Disinformation and misinformation

While the Committee report notes that the broader, societal harms caused by social media platforms were not the focus of this inquiry, Labor members were concerned by the inadequacy of existing responses to the harms caused by the

amplification of mis- and disinformation by social media platforms – particularly in the context of the upcoming Federal election.

Inconsistent Enforcement of Misinformation Policies and the Lack of Enforcement Transparency

The Committee heard significant evidence that social media platforms are inconsistently enforcing their published policies on misinformation and disinformation. Compounding this, the platforms offered little transparency about the enforcement actions taken in response to users who repeatedly share misinformation.

In the context of the upcoming Federal election, it was particularly disturbing that Google could not explain why it was unable to stop misinformation being amplified on YouTube through paid advertising.

When asked about the United Australia Party's then $5 million spend on advertising on 57 YouTube videos since August 2021, of which six were removed for breaching YouTube's COVID-19 misinformation policies, Google's advice was that it had increased investments in its automated machine learning and AI systems to detect misinformation. It is unclear why machine learning and AI systems were necessary to scrutinise the advertising of users who have a lengthy history of distributing misinformation in contravention of the platforms own community guidelines.

Google's evidence was that the removal of these UAP six videos did not trigger their "three-strikes" policy on the sharing of misinformation as "if a number of videos are found to be violative at the same time, they're bundled into one strike in a period of time." Google did not expand on how many videos would need to be removed or the period of time that needs to elapse before triggering a second or third strike.

Google were also unwilling to disclose how many strikes they have issued against the United Australia Party for the sharing of misinformation on YouTube, noting:

> We preserve and balance the privacy of users on our platforms with the need for accountability to our members of parliament. Indeed, in a committee at which we appeared at the end of last year, we provided confidential information to committee members, to members of parliament, about particular strikes that had been issued. That in that way ensures that we

> balance the rights to privacy with the rights and the need for accountability in our policymakers and members of parliament.[1]

Labor members are of the view that this approach does not strike an appropriate balance between privacy and the public interest. In the context of the serious harms caused by misinformation during the COVID-19 pandemic, and the 2022 Federal election, there is a broader public interest in understanding Google's specific approach to the enforcement of its misinformation policies against public figures and political actors.

It is not in the public interest for enforcement action taken against political actors to be a black box that prevents public scrutiny of these actions. Public confidence in the integrity of these matters relies on the public at large, not just the subjects of enforcement action, being able to understand the enforcement actions taken against political actors for identified breaches of the platforms policies. This is particularly so when these platforms are the recipients of tens of millions of dollars in advertising revenues from the subjects of enforcement action.

It does not help that YouTube's problems with mis- and dis-information are well documented. The International Fact-Checking Network, a coalition of 80 international fact-checking groups including RMIT ABC Fact Check, in Australia, sent an open letter to the Chief Executive Officer of YouTube, stating:

> YouTube is allowing its platform to be weaponized by unscrupulous actors to manipulate and exploit others, and to organize and fundraise themselves.[2]

Any steps taken by Google to provide transparency in its enforcement decision making will assist to provide public confidence that this issue is being taken seriously by the company.

Limiting the Reach of Mis- and Disinformation – The Fact Checking Choke Point

The Committee also heard evidence with respect to the fact check checking processes run by the social media platforms. Each of the major platforms gave evidence that once content was identified as misinformation through these fact checking processes, platform specific measures were taken to limit the extent to

---

1 Ms Lucinda Longcroft, Director, Government Affairs and Public Policy, Google Australia and New Zealand, *Committee Hansard*, 20 January 2022, p. 6.

2 International Fact-Checking Network, 'An open letter to YouTube's CEO from the world's fact-checkers', *Poynter*, 12 January 2022, available at: https://www.poynter.org/fact-checking/2022/an-open-letter-to-youtubes-ceo-from-the-worlds-fact-checkers/ (accessed 15 March 2022).

which the reach of this content could be amplified. Despite this, none of the social media platforms gave any commitments about either the expected time frames in which these fact checks would occur, nor the resources that would be allocated to the task. In the context of a month long election campaign, misinformation that took weeks to be fact checked under current processes has the potential to cause significant public harm before the platforms took action.

Given that the risks to social cohesion and democratic integrity from mis- and dis-information are amplified during periods of intense political debate, Labor members were concerned that Meta was unable to provide advice on the time taken to fact check information in the context of the 2022 election.

Meta advised the Committee that is unable even to report on the average time taken to fact check mis- and dis-information as:

> [O]ur fact checkers operate independently it's really up to them to apply their good judgements and their editorial techniques in order to undertake whatever research they need to in order to verify a claim.[3]

Meta was also unable to provide answers on the time taken to fact-check and demote a particularly egregious piece of disinformation from the 2019 election campaign. It is easy to be cynical about the social media platforms intent to improve the timeliness of their fact checking process when they do not appear to be even measuring their own performance in this regard.

Labor Members request that social media platforms set public benchmarks for the performance of their fact checking and misinformation demotion processes in the context of the 2022 Federal Election.

Regulating Misinformation in Australia – Release the Report, Minister!

The inconsistency of the platforms' response to mis- and dis-information has been compounded by the ambiguity of the regulatory context in Australia.

Despite the ACCC Digital Platforms Inquiry recommending an enforceable industry code on disinformation and misinformation in July 2019, the ACMA still hasn't been formally empowered by the government to act in relation to these issues.

Instead, the Digital Industry Group Inc published a voluntary industry code, the Australian Code of Practice on Disinformation and Misinformation on 22 February 2021. Many witnesses who appeared before the committee expressed reservations

[3] Mr Josh Machin, Head of Public Policy, Meta, *Committee Hansard*, 2 March 2022, p. 25.

about the limited obligations and scope for review and enforcement action created under this code.

ACMA provided the Minister for Communications, Urban Infrastructure, Cities and Arts with a report on the impacts of misinformation in Australia, encompassing the impacts of misinformation during the COVID-19 pandemic and in the context of Australian electoral processes, on 30 June 2021. The ACMA also expressed a view about the adequacy of the measures included in the voluntary industry misinformation code in this report.

Despite the salience of this ACMA report, to both the ongoing COVID-19 pandemic and the imminent Federal Election, the Minister has been 'considering' this report for nearly 9 months now.

Labor Members recommend that the Minister for Communications, Urban Infrastructure, Cities and the Arts immediately release the ACMA's misinformation report and respond to the ACMA's recommendations.

The Broader, Societal Harms of Social Media Platforms – Hate Speech and Violent Radicalisation

As outlined in the Committee Report, Australia's existing online safety laws are currently focused on harms directed towards individuals. However, Labor members are of the view that greater attention should be paid to the broader, societal harms caused by social media including particularly harms to our social cohesion and to our democracy. A glaring gap in the existing Australian regulatory regime compared to other nations is dealing with hate speech targeting groups.

This is particularly concerning in the context of this report being handed down at the time of the third anniversary of the Christchurch terrorist atrocity, in which a 28-year-old Australian man entered two mosques in the New Zealand city of Christchurch and murdered 51 people and injured 40.

While the Abhorrent Violent Material legislation introduced after the Christchurch terrorist attack regulates materials that are the product of violent extremism, we have yet to address the online material that normalises hate and radicalises people to commit these acts of real-world violence.

The Committee heard evidence that the current online safety regime was not adequately equipped to deal with online abuse directed at groups of people rather than individuals. It further heard that abuse directed at groups on the basis of their ethnicity and religion – most commonly at people of Muslim and Jewish faith – was a particular problem to which there was little recourse.

The Committee heard from Ms Rita Jabri Markwell of the Australian Muslim Advocacy Network on the impact this type of abuse, and the subsequent failure of government to address it in the wake of the Christchurch attack, saying "it's made us feel really lonely. I don't know how else to describe it. It's kind of like you don't matter".[4]

The status quo was exemplified by Ms Markwell's experience in trying to get former Senator Fraser Anning's account removed from Facebook. The Australian Muslim Advocacy Network successfully brought a vilification action against former Senator Fraser Anning under the Queensland Anti-Discrimination Act. Mr Anning's page had been publishing a constant stream of islamophobia and content promoting the 'great replacement' hate narrative that inspired the Christchurch terrorist and a series of right-wing extremists before him. In response, the tribunal ordered the removal of 141 hate artefacts from Mr Anning's page. Despite this, Facebook did not remove Mr Anning's Facebook page.[5]

The Online Hate Prevention Institute and the Executive Council of Australian Jewry also spoke about working with platforms and ISPs to have hate content, like terrorist manifestos, removed. They told the Committee that they had difficulty getting eSafety and the Australian Communications and Media Authority to assist to remove such content from the internet.[6]

Civil society groups and industry are in rare agreement on the need for action on this issue. In submissions to the inquiry, Reset Australia - a policy think-tank that advocates for technology policy reform - and DIGI - the industry lobby group for major technology platforms including Facebook, Twitter, and Google – have both called for such reforms.

Rest Australia noted in their submission that:

> Expanding the definitions of harms (and risks) addressed in Australia's regulatory framework would better protect Australian communities and society at large. This means tackling mis and disinformation, and explicitly addressing hate speech.[7]

---

4 Ms Rita Jabri-Markwell, Adviser, Australian Muslim Advocacy Network, *Committee Hansard*, 22 December 2021, p. 26.

5 Ms Rita Jabri-Markwell, Adviser, Australian Muslim Advocacy Network, *Committee Hansard*, 22 December 2021, p. 21.

6 Dr Andre Oboler, Chief Executive Officer and Managing Director, Online Hate Prevention Institute, *Committee Hansard*, 22 December 2021, pp 13-14.

7 Reset Australia, *Submission 12*, p. 6.

The platforms too have called for action in their individual capacities. Twitter noted in their submission that:

> Conversely, Australia continues to utilise a very limited definition of hate speech under the Racial Discrimination Act 1975 (Cth) that is limited to race-based speech or behaviour, and does not include a number of the aforementioned categories, including sexual orientation, disability-based, religious-based or gender-based speech.[8]

Mr Josh Machin of Meta stated in evidence to the Committee that:

> We'd encourage Australian policymakers to look at what could be done in relation to hate speech online and offline. State based discrimination laws are obviously intended to provide remedies for particular individuals who can show some disadvantage that they've received on the basis of that discrimination. I think the difficulty with hate speech is that it's not always targeted at individuals.[9]

Three years on from the Christchurch atrocity, and with right-wing extremism a growing concern for Australian security agencies, we need to address online hate speech.

Labor members consider that Australia's online safety framework should be updated to enable action to be taken against group hate speech.

Labor members of the Committee are also of the view that further work needs to be done to improve the government's efforts to share information and collaborate with industry to reduce exposure to terrorist violent and extremist content.

In evidence to the Committee the Department of Home Affairs stressed the importance of "constructive collaboration between government, the community and industry".[10] This is consistent with Recommendation 4.3 of the final report of the Australian Taskforce to Combat Terrorist and Extreme Violent Material Online, which commits government to proactively "share with digital platforms (where legally and operationally feasible) indicators of terrorism, terrorist products and depictions of violent crimes."

Meta, the parent company of Facebook, Instagram, and WhatsApp informed the Committee that:

---

8 Twitter, *Submission 50*, p. 8.

9 Mr Josh Machin, Head of Public Policy, Meta, *Committee Hansard*, 2 March 2022, p. 18.

10 Mr Brendan Dowling, First Assistant Secretary, Digital and Technology Policy Division, Department of Home Affairs, *Committee Hansard*, 1 February 2022, p. 1.

> The Department of Home Affairs has not shared any intelligence specifically about dangerous organisations or individuals with Meta (noting that there are referrals on other issues such as social cohesion and misinformation).[11]

This was confirmed by the Department of Home Affairs. In response to Questions on Notice the Department of Home Affairs noted that it had shared two academic papers, a literature review, and materials publicly available through the Australian Security Intelligence Organisation's Outreach portal.[12]

Labor members of the Committee recommend that the Department of Home Affairs evaluate its practices with respect to notifying social media providers of terrorist violent and extremist content on their platforms, with a view to improving intelligence flows from the Department to social media providers.

Labor members of the Committee recommend that the government immediately release its formal assessment of the actions detailed in the final report of the Australian Taskforce to Combat Terrorist and Extreme Violent Material Online.

The Importance of Multistakeholder Governance Models in Online Safety

As a general observation from the conduct of this inquiry, Labor members wish to underscore the ongoing importance of multistakeholder governance models for all parties with policy responsibility in this space, including governments, regulators and social media platforms.

The Committee heard repeated evidence that the kinds of online harms considered by this inquiry, and the impacts of potential policy and regulatory responses to these harms, manifest differently on different groups within the community.

It is important that groups that have been historically marginalised on the grounds of race, religion, sexuality, gender identity, disability and occupation (including sex workers), are included in multistakeholder governance processes at the dialogue, decision making and implementation stages, to ensure that the impacts of online harms and regulations on these groups are fully appreciated.

This is why Labor moved amendments formalising consultative committee processes at the eSafety Commission as part of the Parliamentary debate on the

---

11 Meta, *Submission 49.1*, p. 26.

12 Department of Home Affairs, *Submission 40.1*, p. 8.

Online Safety Bill, but this remains an ongoing challenge that requires further attention by all actors in this space.

Consolidating the Online safety regulatory framework

Labor members acknowledge the Committee report's suggestion that "the OSA should be reviewed in future to ascertain whether it had made an effect in relation to online safety and to examine whether a single regulatory framework would reduce complexity and confusion for providers."

Labor members note that this is in-line with the 2018 Report of the Statutory Review of the Enhancing Online Safety Act 2015 and the Review of Schedules 5 and 7 to the Broadcasting Services Act 1992 (Online Content Scheme) ('the Briggs review').

The Briggs review suggested that:

> The majority of submitters to these reviews proposed taking Schedules 5 and 7 out of the Broadcasting Services Act and incorporating them within the Enhancing Online Safety Act after a thorough review to clean up and modernize both pieces of legislation, resulting in a single piece of legislation relating to online services and content. I support this approach most strongly and consider that the legislation should be redrawn into an integrated act which deals systemically with the challenge of online safety in its entirety.[13]

And further:

> that there should be a single new fit for purpose and technology-neutral code of practice. This single code would fulfil a wider purpose than the current codes.[14]

Labor members welcome the Committee report's recommendation that Australia's online safety regulation be consolidated in a single regulatory framework. Such a consolidation would improve the effectiveness of the regime for the Australian public and make compliance with Australian regulatory obligations easier for social media platforms, particularly smaller new entrants. However, Labor

---

13 Lynette Briggs AO, *Report of the Statutory Review of the Enhancing Online Safety Act 2015 and the Review of Schedules 5 and 7 to the* Broadcasting Services Act 1992 *(Online Content Scheme),* available at: https://www.infrastructure.gov.au/sites/default/files/briggs-report-stat-review-enhancing-online-safety-act2015.pdf (accessed 18 December 2021), p. 8.

14 Lynette Briggs AO, *Report of the Statutory Review of the Enhancing Online Safety Act 2015 and the Review of Schedules 5 and 7 to the* Broadcasting Services Act 1992 *(Online Content Scheme),* available at: https://www.infrastructure.gov.au/sites/default/files/briggs-report-stat-review-enhancing-online-safety-act2015.pdf (accessed 18 December 2021), p. 14.

members express concern that more than three years after the Briggs' review, the government has failed to follow through on this commitment to consolidation of online safety regulatory regimes, and has instead continued to add additional parallel and overlapping regulatory frameworks.

Coordination between Regulators and Law Enforcement

It has been over three years since Ms Lynelle Briggs AO noted that "even though there are memorandums of understanding between the eSafety Office and all state and territory police services, the Department of Home Affairs' submission to these reviews suggests that there is sufficient uncertainty around these arrangements."[15]

Labor members are concerned that the evidence heard by this inquiry suggests that this recommendation has been ignored.

In the Senate Legal and Constitutional Affairs Legislation Committee, the eSafey Commissioner expressed the view that "it is up to the law enforcement agencies themselves, which I note have billions of dollars in funding, hundreds of thousands of workers to train, local police, and know what resources are available to them."[16]

However, the eSafety Commissioner has failed to update these memoranda, noting in evidence to the Legal and Constitutional Affairs Legislation Committee of the Senate the "Our MOUs don't go to the kinds of conversations and collaboration that we have [with NSW police]."[17]

Labor members recommend that eSafety immediately review its arrangements with state and territory police forces, to prevent further cases from "fall[ing] through the cracks" without receiving services from the responsible Commonwealth agency.

---

15 Lynette Briggs AO, *Report of the Statutory Review of the Enhancing Online Safety Act 2015 and the Review of Schedules 5 and 7 to the* Broadcasting Services Act 1992 *(Online Content Scheme),* available at: https://www.infrastructure.gov.au/sites/default/files/briggs-report-stat-review-enhancing-online-safety-act2015.pdf (accessed 18 December 2021), p. 26.

16 Ms Julie Inman Grant, eSafety Commissioner, Office of the eSafety Commissioner, Inquiry into the Social Media (Anti-Trolling) Bill 2022, Senate Legal and Constitutional Affairs Legislation Committee, *Committee Hansard*, 10 March 2022, p. 18.

17 Ms Julie Inman Grant, eSafety Commissioner, Office of the eSafety Commissioner, Inquiry into the Social Media (Anti-Trolling) Bill 2022, Senate Legal and Constitutional Affairs Legislation Committee, *Committee Hansard*, 10 March 2022, p. 18.

Technology Policy Coordination

Labor members note that the Committee heard repeated evidence about the impact that duplicative and parallel policy development processes were having on the sector. Google told the committee that it was currently participating in nine separate regulatory processes between the Government and industry that address cyber safety and privacy. Meta told the committee that there had been 18 major Government or parliamentary inquiries or consultations impacting digital platforms over the last three years.[18]

Beyond the immediate scope of this inquiry, technology journalist Stilgherrian has publicly tracked dozens of 'cyber related' bills, regulatory instruments, governmental and parliamentary inquiries, and policy consultations at the Commonwealth level in 2021 alone.

While there is clearly a need for significant policy action to address the substantial and evolving harms in this space, it is clear that the lack of process co-ordination is undermining the effectiveness of regulation and creating real costs.

The Committee heard evidence numerous stakeholders, corporate and civil society alike, that this has resulted in:

- An unnecessary resource burden being imposed on stakeholders responding to these processes;
- Inconsistent objectives being pursued by policy makers and regulators, undermining policy effectiveness and complicating compliance; and
- Recommendations of lengthy policy development processes being rejected or forgotten and ultimately never implemented.

Unfortunately, an ad hoc, 'announcement-led' approach in a complex and rapidly evolving area has led to a series of technology policy coordination issues.

Technology Policy Coordination Issues	
Online Safety Act	▪ It took the Morrison government two years and four months to introduce the Online Safety Bill into the Parliament after it was recommended by the Briggs Review in October 2018.

[18] Meta, *Submission 49*, p. 4.

	▪ During this two-year delay, the eSafety Commissioner has highlighted the way that online harms spiked during the pandemic lockdowns across CSAM, cyber bullying and adult cyber abuse.
Age verification	▪ There are currently five parallel and duplicative age/identity verification schemes under development from four different departments and regulators.
Social Media (Anti-Trolling) Bill	▪ The Bill sits outside existing and ongoing state defamation reform process and state Attorneys-General were not consulted prior to the release of the exposure draft. ▪ Esafety Commissioner *'concerned'* and described the marketing of the Bill as an online safety initiative as 'not ideal' and causing confusion among victims. ▪ The PM repeatedly stated in December that the bill would be *'implemented'* or *'dealt with'* by Parliament in February. It has not yet been brought on for debate.
Social Media Online Safety Select Committee Inquiry	▪ Announced by the Prime Minister on 1 December 2021 with a reporting date of 15 February 2022. ▪ As this reporting date was just three weeks after the Online Safety Act came into force, the Committee was unable to evaluate the effectiveness of the new regulatory framework as part of its work.
Digital Platforms Inquiry	▪ More than 2 ½ years after the report was released by the ACCC in July 2019, much of the report is still unimplemented with important recommendations relating to consumer privacy, disinformation, ad restrictions, digital literacy, and consumer choice yet to be actioned.
ACMA Disinformation Code	▪ The ACCC's Digital Platforms Inquiry recommended an enforceable code to counter disinformation in July 2019. It took the government a year and a half to deliver a voluntary opt-in code, designed by the industry body.

	▪ While the ACMA's provided an evaluation of the code to the Minister 9 months ago, it has not been released or acted upon.
Australian Taskforce to Combat Terrorist and Extreme Violent Material Online	▪ Taskforce report outlining measures to be taken by governments and platforms to combat terrorism in the wake of the Christchurch massacre released in June 2019. ▪ Despite committing to a review of effectiveness of the measures agreed to in the report with a view to regulating if further action was required, no assessment of these measures has been released in the more than two years since.
Privacy Act Review & Online Privacy Code	▪ While both the economy wide Privacy Act Review and the Online Privacy Code were recommended in the ACCC Platforms Report in July 2019, these reform processes are currently being undertaken at the same time. ▪ Given the reliance of the Online Privacy Code on the Australian Privacy Principles, any changes to the APPs resulting from the Privacy Act Review would immediately render the Code outdated and inconsistent.
AHRC Human Rights and Technology Final Report	▪ Released in May 2021, the AHRC Human Rights and Technology report was the result of three years of consultation. The report recommended a stronger regulatory regime for the use of AI. ▪ Government has not responded to the report and has indicated that it has no plans to review its voluntary AI ethics framework.

While the Committee heard that there is now a plethora of internal government policy coordination committees on these matters and a series of ad hoc, bilateral engagement forums between regulators, it is clear that these processes are not preventing the emergence of duplicative and inconsistent policy development processes.

The Department of Infrastructure, Transport, Regional Development and Communications informed the committee in response to questions on notice that despite the current volume of activity in this space, the Agency Heads Committee

on Online Safety has only met once since the start of 2021 and the eSafety Advisory Committee has only met three times in that same period.[19]

It is worth comparing the current approach of the Commonwealth to technology policy coordination with that of financial sector regulation. In the financial sector, the Council of Financial Regulators operates as a co-ordinating body with a terms of reference established under a memorandum of cooperation. The CFR charter provides that it is established to pursue a shared objective (the stability of the Australian financial system) and provides a forum to pursue this objective through:

- Identification of important issues and trends;
- Information exchange and coordination where members responsibilities overlap;
- Harmonisation efforts to align reporting and regulatory requirements;
- Coordinating international engagement with peer bodies overseas.

Australian technology policy would benefit from a similar institution (potentially including the OAIC, eSafety, ACMA, ACCC, ACSC, and relevant Government Departments) to assist with the identification of emerging issues, coordination of regulatory processes and alignment of policy objectives.

Labor members note the announcement made near the conclusion of this inquiry that the ACCC, ACMA, the Australian Information Commissioner and the eSafety Commissioner intend to form a Digital Platform Regulators Forum to coordinate their actions. After the extensive evidence of repeated failures of technology policy coordination heard by this committee, this is a welcome development.

However, it must be recognised that the Federal government itself has been the source of the bulk of the policy coordination failures and process duplication in recent years. The success of any technology policy coordination institution will therefore depend on the involvement and buy in of Commonwealth technology policy development agencies, and not just the regulators.

Labor Members recommend that the Government consider the establishment of a Council of Technology Regulators, modelled on the Council of Financial Regulators, to coordinate and align technology policy making in Australia.

---

19 Department of Infrastructure, Transport, Regional Development and Communications, *Submission 44.1*, pp 6-19.

Mr Tim Watts MP

Deputy Chair

Labor Member for Gellibrand

Ms Sharon Claydon MP

Labor Member for Newcastle

Additional comments by Craig Kelly

Social media platforms such Facebook, YouTube, Twitter and TikTok now dominate what was once the 'town square' - the public place available to all-comers, the place where ideas were rigorously debated, impassioned political speeches made, and where discussion on controversial subjects occurred. These social media giants have today in effective privatised the old 'town square'.

The censorship policies and the deplatforming of individuals by these social media giants has and continues to undermine freedom of speech, and present a dangerous threat to our democracy.

The social media giants have used so-called ''fact-checkers'' to falsely portray themselves as the 'arbiters of truth', when in fact these so-called 'fact-checkers' offer little more than an alternate opinion.

The debates and proceedings of the Australian Parliament are protected from censorship by the Bill of Rights 1688, yet You-Tube in particular have shown contempt for Australia's democracy and our nation's laws by censoring the proceedings of the Australian Parliament.

Further, YouTube have threatened Parliamentarians with censorship and deplatforming if they post any proceedings of the Australian Parliament where an opinion was expressed that does not align with You Tube's deceptively labelled ''community standards''.

Open and free elections in Australia are threatened by social media giants censoring and deplatforming the political arguments and policy platforms of election candidates and registered political parties.

The censorship, secret shadow-banning and deplatforming of candidates or political parties by social media giants acting as 'platforms' (and not as publishers) during an election campaign, places that candidate at a competitive disadvantage

to other candidates. This amounts direct foreign interference in Australian elections and poses a grave threat to our nation's democracy.

Our history teaches us that there is no group of experts adequate enough or wise enough to operate without scrutiny or criticism. Yet during the COVID period the social media giants engaged in unprecedented censorship and deplatforming of highly qualified expert medical opinion that sought to scrutinise or criticise the narrative and the press releases issued by the COVID vaccine sellers.

In particular, the social media giants relentlessly censored expert medical opinion and scientific evidence (and deplatformed medical doctors for expressing opinions) which questioned the narrative that early treatment of COVID with existing repurposed off-patent medicines such as HCQ and Ivermectin (when combined with zinc and an antibiotic and administered within the first few days after infection) was not only ineffective but dangerous and needed to be banned.

With the passage of time, and the vast volume of scientific evidence that has since accumulated, this previously censored expert medical opinion and the deplatformed medical doctors have been proven correct.

By their conduct of deplatforming medical doctors and censoring expert medical opinion & the scientific evidence on early COVID treatments, the social media giants, You-Tube, FaceBook & Twitter have prevented free and open debate on effective early treatment, contributing (in the opinion of highly qualified medical experts) to countless additional hospitalisations and deaths. There is no other conclusion that You-Tube, FaceBook and Twitter have ''blood on their hands'' as a result of their censorship and suppression of debate.

Recommendations

A. Penalties under the s327 of Electoral Act 1918 include imprisonment for hindering or interfering with a persons free exercise of a political right of liberty. Consistent with the criminal penalties in the existing laws for interference in Australian elections, senior management of the social media platforms should likewise face imprisonment for interfering in Australian elections by engaging in any conduct that censors, shadow-bans or deplatforms any candidate for political office engaging in otherwise lawful political debate.

In short, if it's lawful to say in the town square at a public meeting, it should be unlawful for foreign social media giants to censor it. And to demonstrate that as a nation that we are serious about protecting free speech and our democracy, criminal penalties are justified.

B. The Parliament , without further delay immediately pass the private members Bill, Social Media (Protecting Australians from Censorship) Bill 2022, introduced by the Member for Dawson.

This bill prohibits large foreign social media services from de-platforming or censoring lawful content by Members of Parliament, election candidates, registered political parties, journalists and media organisations on their platforms within Australia.

The bill also prohibits large foreign social media services from censoring lawful philosophical (including political) discourse on their platforms within Australia.

The bill tasks the Australian Communications and Media Authority to oversee the administration of this proposed law and the issuance of notices and fines to foreign social media services.

The bill supports the right to freedom of speech within Australia. The bill also seeks to minimise opportunities for foreign social media services to influence elections and political discourse in Australia.

Further comments

Throughout the enquiry the Deputy Chair made repeated assertions alleging that the United Australia Party had posted 'misinformation' online that needed to be censored. As leader of the UAP, I specifically reject these assertions. Alternative opinions that vary from ''the narrative'' are not 'misinformation' but alternate opinions.

A wise Canadian judge has recently commented on the use of the slur 'misinformation' noting,

"And is 'misinformation' even a real word ? Or has it become a crass, self-serving tool to pre-empt scrutiny and discredit your opponent ..... A childish - but sinister - way of saying, 'You're so wrong. I don't even have to explain why you're wrong'".

Mr Craig Kelly MP
Member
UAP Member for Hughes