Submission to the Select Committee on Social Media and Online Safety

on Inquiry into

Social Media and Online Safety

12 January 2022

DIGITAL
RIGHTS
WATCH

Overview

Digital Rights Watch (DRW) welcomes the opportunity to submit comments to the House Select Committee on Social Media and Online Safety regarding the Parliamentary Inquiry into Social Media and Online Safety. DRW has been an active participant in the public consultations related to the Online Safety Act over the course of 2021, including:

- Submission on the proposed Online Safety Bill[1]
- Submission to the Senate Inquiry into the Online Safety Bill[2]
- Participation in the Public Hearing for the Senate Inquiry into the Online Safety Bill[3]
- Submission on the Restricted Access Systems Discussion Paper[4]
- Submission on the draft Basic Online Safety Expectations[5]
- Submission on the draft Restricted Access Systems Declaration[6]

DRW has also participated in other consultations relevant to social media and online safety including:

- Submission to the Privacy Act Review Discussion Paper[7]
- Submission to the inquiry into the Abhorrent Violent Material Act[8]
- Submission on the draft Online Privacy Bill[9]

The contents of these submissions are of direct relevance to the Committee's Terms of Reference for the Inquiry into Social Media and Online Safety, and we encourage the consideration of them as part of this Inquiry.

[1] Digital Rights Watch, Submission to the Department of Infrastructure, Transport, Regional Development and Communication on the proposed Online Safety Bill 2020, 14 February 2021, available at: https://digitalrightswatch.org.au/2021/02/18/submission-the-online-safety-bill/

[2] Digital Rights Watch, Submission to the Senate Standing Committees on Environment and Communications on the proposed Online Safety Bill 2020, 2 March 2021. Submission number 27, available at: https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Environment_and_Communications/OnlineSafety/Submissions

[3] Environment and Communications Legislation Committee Public Hearing, 5 March 2021, transcript available at: https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Environment_and_Communications/OnlineSafety/Public_Hearings

[4] Digital Rights Watch, Submission to the eSafety Commission on the Discussion Paper on the Restricted Access Systems Declaration, 17 September 2021, available at: https://digitalrightswatch.org.au/2021/09/21/submission-restricted-access-system/

[5] Digital Rights Watch and Global Partners Digital, Submission to the eSafety Commission on the draft Online Safety (Basic Online Safety Expectations) Determination 2021, 4 November 2021, available at: https://digitalrightswatch.org.au/2021/11/04/submission-draft-basic-online-safety-expectations/

[6] Digital Rights Watch, Submission to the eSafety Commission on the draft Restricted Access Systems Declaration 2021, 23 November 2021, available at: https://digitalrightswatch.org.au/2021/11/25/submission-draft-restricted-access-systems-declaration/

[7] Digital Rights Watch, Submission to the Attorney-General on the Discussion Paper regarding the Review of the Privacy Act 1988, 10 January 2022. Available at: https://digitalrightswatch.org.au/2022/01/11/submission-privacy-act-review-discussion-paper/

[8] Group submission to the Parliamentary Joint Committee on Law Enforcement in relation to the Abhorrent Violent Material Act, October 2021, available at: https://digitalrightswatch.org.au/2021/10/29/submission-abhorrent-violent-material-act/

[9] Digital Rights Watch, Submission to the Attorney-General on the proposed Privacy Legislation Amendment (Enhancing Online Privacy and Other Measures) Bill 2021, 6 December 2021, available at: https://digitalrightswatch.org.au/2021/12/07/submission-online-privacy-bill/

About Digital Rights Watch

Digital Rights Watch is a charity organisation founded in 2016 whose mission is to ensure that people in Australia are equipped, empowered and enabled to uphold their digital rights. We stand for Privacy, Democracy, Fairness & Freedom in a digital age. We believe that digital rights are human rights which see their expression online. We educate, campaign, and advocate for a digital environment where individuals have the power to maintain their human rights.[10]

General remarks

At DRW we recognise that there are unique and complex challenges posed by the ubiquitous nature of the internet and digital platforms, and the legitimate interest of the Australian government to promote safer online services to individuals across Australia.

Harms caused or exacerbated by social media and other digital platforms require significant scrutiny. The Australian and international digital rights community has long called into question the harmful practices of digital platforms, suggesting improved regulation of digital platforms and online services, including improved protections for privacy and digital security for individuals. The revelations contained in the leak of the Facebook Papers are a clear example of some of the harms that social media companies knowingly create or worsen, and highlight many of the concerns that those in the digital rights community have been sounding the alarm on for years.[11]

Over the course of the past 18 months, the Australian government has directed significant attention towards the regulation of Big Tech, with a particular emphasis on online safety. This culminated in the introduction of and ongoing consultations surrounding the *Online Safety Act,* the *Online Privacy Bill,* and the *Social Media (Anti-Trolling) Bill.*

At DRW we are committed to improving online safety for everyone. As such, we have actively participated in the public debate, including government consultations, expert roundtables, and community engagement. We believe that the inclusion of civil society, technology experts, and affected communities in the public debate is essential in order to reach a functional and effective approach to harm reduction. We are invested in achieving the balance between strong regulation and individual rights and freedoms. Unfortunately, the development of the approach to online safety in Australia, including the approach to community engagement and public consultation suggests that the government is not upholding its duty to multistakeholderism. The rushed public consultation process in the development of the Online Safety Act,[12] and the overwhelming number of consultations concentrated over the festive season,[13] are key examples that demonstrate the flawed approach to meaningfully engaging with civil society, industry, and other interested parties.

[10] Learn more about our work on our website: https://digitalrightswatch.org.au/

[11] https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers/

[12] We note that over 300 submissions were made to the initial public consultation Insert article here that references the rushed process.

[13] https://www.innovationaus.com/extremely-difficult-govts-summer-submission-rush/

The approach to online safety must always take into account the complexities of modern life, norms, and digital technologies. When the focus of online safety is solely on the symptoms—namely online abuse and harassment, misinformation, and defamation—without also considering the underlying business models, technological realities, legislative landscape, and social norms, the government risks *creating* additional online harms, in its pursuit of mitigation. This is made worse by failing to listen to, or refusing to incorporate, feedback from the most affected communities.

Within this submission we focus on the following key points, which we encourage the committee to consider over the course of the inquiry:

1. **Address the causes, not just the symptoms**
2. **Online anonymity and pseudonymity are important, and warrant protection**
3. **Age verification proposals create more harm than good**
4. **Automated content moderation is not a simple fix**
5. **Robust digital security keeps us all safer in a connected world**

We urge the Committee to take care to avoid conflating surveillance with safety. Increasing data collection, proactive monitoring, undermining security measures such as encryption, and emphasising policing of online spaces may create the illusion of safety for *some,* but it will not result in safer online experiences for *all* Australians. This is because systems of surveillance disproportionately target and penalise individuals and communities who are already under-represented, overly-surveilled, marginalised or have a history of oppression. It is essential that intersections of gender, class and race are part of any consideration of online safety.

We welcome the opportunity to meet with the Committee, to appear at a public hearing, or to otherwise offer our perspective and expertise to the Inquiry.

1. Address the causes, not just the symptoms

> *Term of Reference:*
> *(g) actions being pursued by the Government to keep Australians safe online;*

At DRW we have welcomed the findings of the ACCC Digital Platforms inquiry which made extensive recommendations regarding the need for a data protection framework and improved protections for privacy in order to protect Australian consumers.[14] We were concerned that the first actionable outcome of the extensive ACCC inquiry did not focus on addressing the most pressing systematic data collection and exploitation models that digital platforms thrive on, rather, it sought to make sure that news corporations further benefit from

[14] The ACCC Digital platforms final report provides several recommendations on how to strengthen the rights of consumers in the digital space, including stronger privacy protections and data rights: https://www.accc.gov.au/publications/digital-platforms-inquiry-final-report

them by way of the News Media Bargaining Code. This is a prime example of legislation which seeks to address a symptom, rather than the underlying business model of surveillance capitalism which should be the focus of policy makers.

The Code as drafted perpetuates and capitalises on existing advertising-based models which aggregate massive amounts of personal data for profit.[15] It is well established that this contributes to dysfunctional and harmful online spaces in a variety of ways. The Code seeks to course-correct this behavior to redistribute capital among news providers, rather than to minimise the practice of collecting and reselling our behavior patterns and personal information online altogether.

We reiterate the need to follow through with the remaining recommendations of the ACCC Digital Platforms Inquiry.

Many aspects of the Online Safety Act and the proposed Online Privacy also fall into the trap of seeking to address the symptoms of harmful business models, rather than getting to the root cause. For instance, the Online Content Scheme contained in the Online Safety Act incentivises increased content moderation and proactive surveillance upon digital platforms. In doing so, it seeks to address the harm caused by certain forms of content online, but promotes an approach that exacerbates, rather than challenges, the data-hungry and surveillance-driven models of Big Tech.

Similarly, the proposed Online Privacy Bill seeks to reduce harm to children and young people on social media platforms. By pushing for age verification in order to afford one part of the population additional protections, it creates additional privacy and security risks for everyone, as well as feeding directly into the excessive data-collection practices of social media platforms. The intention behind such approaches is understandable, as they may be considered to be ‘easy wins’ in the face of a seemingly insurmountable task of tackling the systemic issues entwined in surveillance capitalism. However, an overly simplistic approach will not yield the desired regulatory result all the while threatening to make the situation actively worse by *increasing* the risk of downstream harms.

We suggest that the Committee consider the following over the course of the inquiry:

- The virality models which underpin how widely and quickly content is shared to a large audience. While the ability to share content remains valuable for people to connect, build communities, and deliver important messages, it is worth interrogating how limiting the virality of content can assist in minimising related online harms. For instance, limiting cross-platform sharing of content, placing safeguards upon such functionality. The share button, for example, was acknowledged by Facebook as being a cause of harm in the Facebook Papers.
- The engagement algorithms, in particular the prioritisation of provocative, inflammatory or otherwise emotive content in order to elicit the maximum amount of engagement. Engagement algorithms which take individuals down ‘rabbit holes’,

[15] Internet Health Report from Mozilla, ‘The Good, The Bad and The Ugly Sides of Data Tracking’, April 2018: https://internethealthreport.org/2018/the-good-the-bad-and-the-ugly-sides-of-data-tracking/

showing progressively more and more extreme content should also be interrogated, as well as how to effectively limit these phenomena.

- The concentration of power and dominance of a handful of Big Tech companies on the Internet. One of the dilemmas of regulating digital platforms is that it can result in the loss of smaller, independent platforms, who struggle to meet the regulatory requirements that have been designed with Big Tech companies in mind. A pluralistic internet requires a pluralistic approach to regulation. Otherwise, we risk regulating smaller entities out of existence that might otherwise be a source of accountability and limit on the power of Big Tech.
- Robust regulation of how companies can collect, use and disclose personal information would go a long way to mitigate many downstream privacy-related harms including hyperpersonalised targeted content or marketing. Focusing on system-wide issues such as privacy and AdTech have more potential for impact, as opposed to narrow focuses such as defamation on social media.

2. Anonymity and pseudonymity online are important and warrant protection

Terms of Reference:
(b) evidence of: (iii) existing identity verification and age assurance policies and practices and the extent to which they are being enforced, and
(c) the effectiveness, take-up and impact of industry measures, including safety features, controls, protections and settings, to keep Australians, particularly children, safe online

Anonymity and pseudonymity online are vital for a variety of reasons.

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.

Given the Inquiry's focus on safety, we wish to emphasise that the use of pseudonyms is a method many Australians use to protect themselves on social media and other digital platforms. Over the course of our advocacy work, DRW has heard from numerous members of the public with regard to how they use pseudonymity online as a safety mechanism.

People who use pseudonyms online do so for a variety of reasons. For example:

- People from marginalised groups—including those from the LGBTQ+ community, disabled Australians, those from ethnic minorities—to build communities online while managing risk to their health, safety, reputation or well being
- People seeking health information or support for stigmatised conditions
- Victim-survivors of domestic violence
- Whistleblowers revealing information about institutional corruption, as well as activists and human rights lawyers working on sensitive topics
- Sex workers building professional networks which provide social support, health and safety information
- Anyone in a public-facing role, such as social or youth workers, case managers, and lawyers, who wish to be able to maintain an online life without being tracked down or contacted by clients or those they work with
- Individuals working in the public sector, who are generally not permitted to make public comments regarding politics or government positions, who wish to participate in online political discussion without jeopardising their role or being perceived to speak on behalf of a government agency

In 2021, DRW co-hosted an expert roundtable to explore how and why anonymity and pseudonymity online is so important. It includes Dr David Kaye, the Former United Nations Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Dr Emily van der Nagel, a researcher and expert in social media identities, platforms and cultures with a focus on anonymity and pseudonymity, among others. One of the key themes which emerged from the discussion is that research has repeatedly shown removing anonymity is not an effective method for reducing harm in online spaces. We strongly suggest the Committee consider this roundtable over the course of the inquiry.[16]

3. Age verification proposals create more harm than good

> *Term of Reference: (b) evidence of: (iii) existing identity verification and age assurance policies and practices and the extent to which they are being enforced*

Age verification (AV) or "age assurance" has been proposed as way to increase online safety by restricting access to certain kinds of material (under the eSafety Restricted Access Systems (RAS) Declaration and the roadmap for AV for access to online pornography), as well as to differentiate the degree of privacy protections (under the proposed Online Privacy Bill).

DRW is deeply concerned about any proposal which requires individuals to verify their age or identity in order to use a digital platform or other online service.

[16] Digital Rights Watch and Twitter, 'Online anonymity and pseudonymity: why it matters', Expert Roundtable Discussion, November 2021. Available at: https://www.youtube.com/watch?v=c_g_hXCW1oY&t=1s

Our concerns can be summarised by:

1) **Age verification is privacy-invasive, which undermines the objective of reducing online harm**
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 conduct age verification creates additional privacy and security risk, which in-turn can exacerbate online harms.

2) **Age verification achieves nothing to change the surveillance-based business models underpinning social media, nor the harms that arise from them**
Many of the harms caused by social media are a result of data-extractive, surveillance-based business models. These models rely on the collection of immense amounts of information in order to be able to target us individually, conduct hyper personalisation, and to shape, curate and manipulate what we are exposed to online. It is clear that these practices do indeed cause significant harm to individuals, especially children. Age verification does nothing to combat these harmful business models, and in fact may cause additional privacy-related harms in the long run, as it relies on collection of additional information which works in *favour* of data-hungry social media platforms.

Approaches to age verification and their privacy and security pitfalls

Age Verification and Restricted Access Systems have been considered in the past but have failed to be implemented due to their overreach, blunt approach, unreasonable impact upon individual's privacy, and the creation of adverse digital security risks.

We are concerned that any of the existing approaches to implementing AV/RAS will require the provision of personal information that goes well beyond proof-of-age. There are significant, if not insurmountable, challenges to implementing age verification in a way that is both effective, as well as minimising privacy and security risk.

Mandatory AV is likely to act as a deterrent for many adults accessing legal content, and may prompt people of all ages toward less safe and secure internet services in order to circumnavigate providing personal information. Further, we remain concerned that many of the current approaches to AV are relatively easily bypassed, for example, by use of a Virtual Private Network (VPN).

The combination of these factors are likely to result in a system which is unduly invasive in data collection, creates new privacy and security risks by holding information on individuals, and yet is unlikely to be effective at preventing people under the age of 18 from accessing restricted content. We are therefore concerned that the outcome will be a system that is not simply ineffective but actively harmful.

There are many technological approaches to age verification. We have identified and grouped the typical approaches below, including a high level overview of our concerns as they relate to privacy, security and digital rights.

1) **A requirement to provide identity documents to the service/platform that hosts the content, or to a third party service, either specifically for age verification, or more broadly as part of identity verification.**

It was only in March 2021 that the House of Representatives Standing Committee on Social Policy and Legal Affairs recommended that in order to have a social media account, individuals should be "required by law to identify themselves to a platform using 100 points of identification, in the same way a person must provide identification for a mobile phone account," as a measure to reduce online abuse.[17] The provision of government identity documents or biometric information to social media platforms was also suggested in August 2021 by the UK Children's Commissioner as a method to restrict access to online pornography.[18]

We are deeply concerned by these proposals, and the impact that such an approach would have upon individuals' right to privacy, their ability to remain anonymous online, and the security of their identity. **We strongly recommend that the RAS Declaration does not allow for any requirement for individuals to provide government-issued identity documents to content providers or digital platforms.**

The risk of identity theft in the event of a data breach whereby personal information is inappropriately or unlawfully accessed and leaked is significant. If the RAS were to require sites hosting sexually-explicit content to collect and hold identifying documentation, it is likely that they would become targets for malicious actors. We remind the Commission of the leak of 30 million accounts when the adultery site, Ashley Madison, was hacked in 2015. The resulting harm caused by such sensitive information being inappropriately-accessed included several deaths by sucide.[19]

We also wish to emphasise the importance of maintaining the ability of individuals to be anonymous online. The suggestion that reducing anonymity would inherently reduce online harms is misguided. In fact, many vulnerable groups including victim-survivors of family violence rely on anonymity online to maintain their safety.[20] As such, any RAS or AV regime must not undermine the ability for people to be anonymous online, and must not require people to provide government-issued identity documents to digital platforms.

[17] 'Inquiry into family, domestic and sexual violence,' *House of Representatives Standing Committee on Social Policy and Legal Affairs*, March 2021, recommendation 30. Available at: https://parlinfo.aph.gov.au/parlInfo/download/committees/reportrep/024577/toc_pdf/Inquiryintofamily,domesticandsexualviolence.pdf;fileType=application%2Fpdf

[18] 'Social Media companies to be told to introduce tough age checks using passports or fingerprint analysis,' *The Telegraph,* August 2021. Available at: https://www.telegraph.co.uk/news/2021/08/30/social-media-companies-told-introduce-tough-age-checks-using/

[19] 'Ashley Madison suicides over web attack,' *BBC,* August 2015. Available at: https://www.bbc.com/news/technology-34044506

[20] See: 'Why Anonymity is Important,' *Digital Rights Watch,* April 2021. Available at: https://digitalrightswatch.org.au/2021/04/30/explainer-anonymity-online-is-important/

With regard to the prospect of using a third-party age-verification service, we wish to emphasise that there should be no information-sharing between the site hosting the restricted content, and the third party providing the age check. For instance, the site providing the restricted content should not be able to access any identification details or know who the person is, and the age verification service should equally not know which site the individual is trying to access, only that age verification is required.

Further, there should never be retention of age-verification data, including metadata logs. If this information were to be retained, it could remain possible to trace or link an identity to their online pornography-viewing habits and preferences, as well as any other 'age-inappropriate material' they may have accessed, which could reveal details about their sexual health or sexual practices. This is an invasion of privacy. Once an individual's age has been verified and they have been granted access, all records of the transaction should be permanently destroyed. There should be no way to retroactively link an individual's identity to the content they have accessed.

2) Verification of age based on user information being cross-checked in other databases that incorporate age-related information.

This approach generally relies on identity or age being validated against verification of another dataset in order to corroborate the information provided by an individual, such as the electoral roll, credit records, or drivers license databases. For example, Equifax suggested in 2019 that "age verification could involve confirmation that a user is listed on the Commonwealth electoral roll or has credit reporting information retained on Equifax's consumer credit bureau, either of which indicates that the user is aged 18 years or above."[21]

However, absence from any of these datasets does not necessarily mean that the individual in question is under the age of 18. We note that the vast majority of adults are either enrolled to vote (96%) and Equifax has estimated that 18 million Australian adults are listed on the customer credit bureau. Nonetheless, this still does not mean that an individual who is not in a database is by default under the age of 18.

We also have strong concerns about the process of cross-referencing information about individuals in datasets controlled by governments at various levels as well as the private sector, and the possibility of inappropriate information sharing or linking across disparate datasets.

Finally, we do not believe that this approach would meet community expectations regarding the use of personal information. When individuals enrol to vote, register for a drivers license, or use a credit card, they have not provided their personal information for the purpose of validating access to restricted material online.

[21] 'Inquiry into age verification for online wagering and online pornography,' *Standing Committee on Social Policy and Legal Affairs,* 2019-2020. Section 2.103. Available at: https://www.aph.gov.au/Parliamentary_Business/Committees/House/Social_Policy_and_Legal_Affairs/Onlineageverification/Report/section?id=committees%2Freportrep%2F024436%2F72614

3) Use of biometric software, either by way of facial recognition or 'age estimation software' that uses photos, videos, or a live stream to estimate age.

The use of facial recognition technology to verify an individual's age by means of checking their identity against a government-issued identity document represents a significant and disproportionate invasion of privacy, and as such, is not an appropriate approach to restricting access to any online content.

In the 2019 'Protecting the Age of Innocence' inquiry, the Department of Home Affairs suggested the use of facial recognition technology by way of its Facial Verification Service (FVS), which it proposed could then be cross-checked with other identity documentation that Home Affairs already holds.[22] However, this is subject to the passage of the *Identity Matching Services Bill 2019,* which we note has received significant public backlash and criticism from privacy and security experts.

The prospect of the Department of Home Affairs utilising facial recognition for the purpose of regulating access to online pornography and other 'age inappropriate material' is unacceptable. No government department, but especially not the one which also contains policing and intelligence agencies within its profile, should be able to associate an individual's biometric data with their sensitive online habits.

We also note that current facial recognition software still exhibits racial and gendered biases, and that by relying on such technology, a RAS may unreasonably prevent an individual who is over the age of 18 from accessing content online, should their face not be recognised by the facial recognition system. By contrast, age estimation software may offer a less privacy-invasive solution, on the condition that no personal information (including biometric information) is collected or retained. However, we would note that the accuracy of age estimation software is questionable at best, and therefore may result in an unacceptable margin of error.

4) Age screening based on requiring users to self-declare, such as through stating their date of birth or ticking a box to state they are over the age of 18.

Many age-restricted online content already employs this method of age restriction, such as when accessing online stores which sell alcohol. While this approach is the least invasive and presents the smallest privacy and security risk, its efficacy is also minimal.

[22] 'Inquiry into age verification for online wagering and online pornography,' *Standing Committee on Social Policy and Legal Affairs,* 2019-2020. Section 2.111. Available at: https://www.aph.gov.au/Parliamentary_Business/Committees/House/Social_Policy_and_Legal_Affairs/Onlineageverification/Report/section?id=committees%2Freportrep%2F024436%2F72614

4. Automated content moderation is not a simple fix

Term of Reference: (b) evidence of: (ii) the extent to which algorithms used by social media platforms permit, increase or reduce online harms to Australians;

DRW has raised concerns regarding approaches to online safety which prioritise or incentivise proactive, automated removal or restriction of online material, otherwise referred to as content moderation. The Online Safety Act, and the accompanying Basic Online Safety Expectations currently use these approaches.

We wish to emphasise that consideration of content moderation cannot simply concern itself with the harms it may prevent by removing content deemed to be harmful. It must also consider the impacts and possible harms that arise when content is wrongly or unreasonably removed or restricted, including when individuals lose access to their online accounts as a result. For example:

- loss of income, livelihoods, and ability to form communities for online content creators, including those in the creative and adult industries
- unreasonable encroachment upon individuals' freedom of expression
- the normative impact upon society as a result of overly restrictive or regressive content moderation rules
- the perpetuation of existing discrimination or biases against already marginalised or overly-policed groups, including but not limited to Black and Indigenous people, People of Colour, disabled people, and the LGBTQ+ community
- the broad, political impact of removing or suppressing content from activists and organisers online working on social movements or other causes, including those sharing abhorrent violent content to shine the light on human rights abuses or excessive use of violent by state actors

Given the scale of online content, digital platforms and other online service providers generally turn to automated processes, including AI, to determine which content is or is not harmful. Larger companies tend to develop their own bespoke tools, whereas smaller companies often purchase or license generic tools for adaptation to their platforms. The risk of encouraging or mandating the use of automated content moderation is that it incentivises platforms toward blanket censorship, and will detect and remove content that is not actually unlawful or harmful in a particular context.

Automated content moderation has been shown to be more technologically effective for video and images, but does not perform well when met with text, audio or other mixed formats. While automated processes have had some success in relation to content moderation with some types of images, such as the ability to scan for copies of images that have already been identified by humans as constituting child sexual abuse and exploitation, other visual-based forms of content moderation on popular social media sites has caused additional harm by disproportionately remove some content over others, penalising Black, Indigneous, fat, and LGBTQ+ people.[23] As experience with the controversial SESTA/FOSTA in the US demonstrated, some platforms will default to blanket removal of all sexual content to avoid penalty rather than deal with the harder task of determining which content is actually harmful.

Cyber-bullying, cyber abuse, and online hate speech, and material which promotes abhorrent violent conduct, may include a mixture of audio, visual and text content and have not been proven to be effectively moderated by automated means. In 2018, Zuckerberg said it's "easier to detect a nipple than hate speech with AI."[24] Automated processes for the detection of such material thus rely on a combination of natural language processing, image recognition and contextual knowledge-mapping for detection, technologies which, at present, are somewhat limited. For example, a recent survey of machine learning techniques for cyber bullying detection on Twitter demonstrated a huge variation in accuracy of different models, which ranged from 30 to 80 percent.[25]

We recommend that any legislative regime avoid incentivising or relying on the use of primarily automated solutions to identify and remove online content, regardless of content category.

5. Robust digital security keeps us all safer in a connected world

Term of Reference:
(g) actions being pursued by the Government to keep Australians safe online;

Without strong digital security such as encryption, the safety of all Australians is undermined.

DRW has previously expressed concern regarding attempts to undermine encryption. For example, the BOSE as currently defined could be made to compromise secure tools and technologies regardless of their overall merit if they somehow impede or prevent investigations by digital platforms into the content defined in Section 46. **We recommend that the Bill is amended to affirm the need for strong encryption and prohibit any interference of the powers prescribed with encrypted tools and technologies.**

The eSafety Commissioner has already publicly argued against end-to-end encryption, saying that it "will make investigations into online child sexual abuse more difficult."[26] While

[23] The algorithms that detect hate speech online are biased against black people: https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter
Facebook repeatedly bans Indigenous activists: https://onlinecensorship.org/content/infographics
Instagram photo censorship: https://www.theguardian.com/technology/2020/oct/20/instagram-censored-one-of-these-photos-but-not-the-other-we-must-ask-why

[24] 'Zuckerberg: It's easier to detect a nipple than hate speech with AI,' Venture Beat, April 2018. Available at: https://venturebeat.com/2018/04/25/zuckerberg-its-easier-to-detect-a-nipple-than-hate-speech-with-ai/

[25] Amgad Muneer and Suliman Mohamad Fati, "A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter", 12(11) Future Internet (October 2020), available at: https://www.mdpi.com/1999-5903/12/11/187

[26] https://www.esafety.gov.au/about-us/blog/end-end-encryption-challenging-quest-for-balance

encryption may impede such investigations, it also provides everyone with digital security, and protects everyone from arbitrary surveillance by malicious actors and cybercrime (e.g. identity theft). Further, it protects the privacy of victims of domestic violence, confidential sources of journalists, safety of political dissidents and all activists, lawyers, and reporters. Claiming that encryption exacerbates harm to children is unproven, and strengthens a regressive surveillance agenda at the expense of everyone's digital security. It is essential that compliance with this Bill does not create a way to compel providers to restrict or weaken their use and application of encryption across their platforms.

DRW have been particularly concerned with the framing of encryption as an inhibitor to safety, which runs counter to the established consensus in the cybersecurity industry who view encryption as vital to *facilitate* safety. Encryption is essential for all businesses, individuals, and digital security at a national level. Encryption facilitates the security of our online activities; protecting data from potential cybercriminals, enabling secure online transactions, and maintaining the privacy and security of our online communications, including those of children. For example, encryption plays a crucial role in preventing malicious actors from accessing networked devices, including tapping into users' webcams or baby monitors.

Any weakening of encryption would undermine the security of Australians' services, jeopardizing the safety of the millions of people who rely on them each day. Prompting services to weaken, undermine, or otherwise bypass encryption threatens the digital security of Australia at a national level by introducing security weaknesses into everyday services used by Australians. Further, encryption is essential for the protection of vulnerable groups, including LGBTQ+ persons[27] and survivors of domestic violence, who rely on encryption to protect the sharing of information about safe relocation, the integrity of digital evidence, and to guard against unauthorised access to survivors' details or communications.[28]

Contact

Samantha Floreani | Program Lead | Digital Rights Watch | samantha@digitalrightswatch.org.au

[27] LGBT Tech & ISOC, Encryption - Essential for the LGBTQ+ Community, available at: https://www.lgbttech.org/post/2019/11/22/lgbt-tech-release-encryption-one-sheet

[28] ISOC, Understanding Encryption Fact Sheet: The Connections to Survivor Safety, (2020) available at: https://www.internetsociety.org/resources/doc/2020/understanding-encryption-the-connections-to-survivor-safety/