Stretching the spectrum of online safety

Submission to the new Online Safety Bill

Submission

Jordan Guiao, Research Fellow

Centre for Responsible Technology

February 2021

About The Australia Institute

The Australia Institute is an independent public policy think tank based in Canberra. It is funded by donations from philanthropic trusts and individuals and commissioned research. We barrack for ideas, not political parties or candidates. Since its launch in 1994, the Institute has carried out highly influential research on a broad range of economic, social and environmental issues.

About the Centre for Responsible Technology

The Australia Institute established the Centre for Responsible Technology to give people greater influence over the way technology is rapidly changing our world. The Centre will collaborate with academics, activists, civil society and business to shape policy and practice around network technology by raising public awareness about the broader impacts and implications of data-driven change and advocating policies that promote the common good.

Our philosophy

As we begin the 21st century, new dilemmas confront our society and our planet. Unprecedented levels of consumption co-exist with extreme poverty. Through new technology we are more connected than we have ever been, yet civic engagement is declining. Environmental neglect continues despite heightened ecological awareness. A better balance is urgently needed.

The Australia Institute's directors, staff and supporters represent a broad range of views and priorities. What unites us is a belief that through a combination of research and creativity we can promote new solutions and ways of thinking.

Our purpose – 'Research that matters'

The Institute publishes research that contributes to a more just, sustainable and peaceful society. Our goal is to gather, interpret and communicate evidence in order to both diagnose the problems we face and propose new solutions to tackle them.

The Institute is wholly independent and not affiliated with any other organisation. Donations to its Research Fund are tax deductible for the donor. Anyone wishing to donate can do so via the website at https://www.tai.org.au or by calling the Institute on 02 6130 0530. Our secure and user-friendly website allows donors to make either one-off or regular monthly donations and we encourage everyone who can to donate in this way as it assists our research in the most significant manner.

Level 1, Endeavour House, 1 Franklin St
Canberra, ACT 2601
Tel: (02) 61300530
Email: mail@tai.org.au
Website: www.tai.org.au
ISSN: 1836-9014

Summary

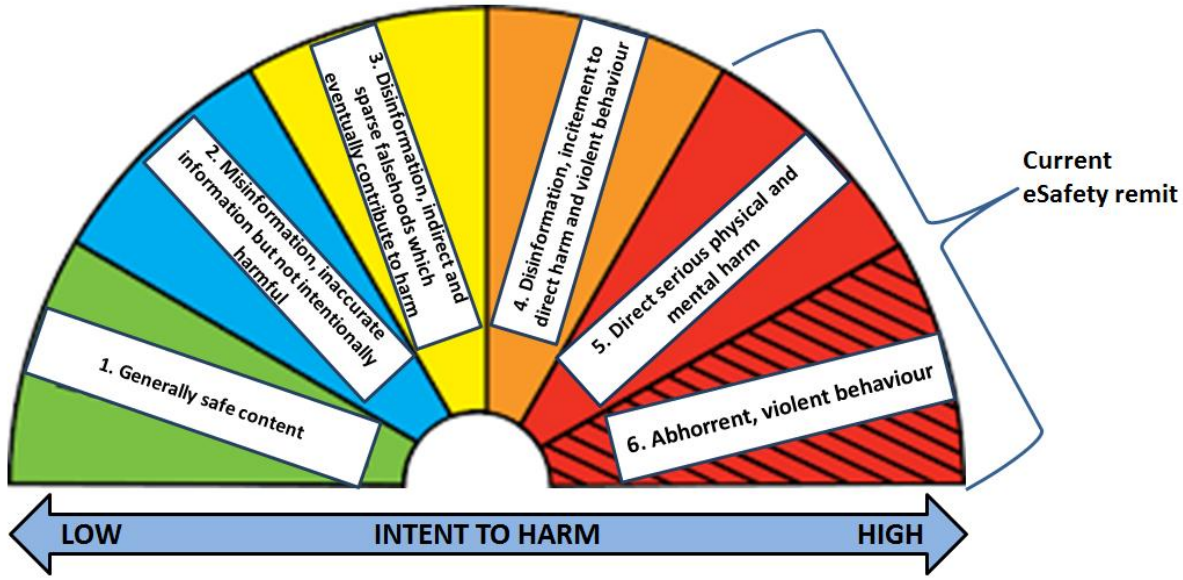
The draft update to the Online Safety Bill recognises the need for legislation to keep pace with the increased harms which are now part of everyday reality for Australians online.

The updates to the draft Bill empowers the eSafety Commission with some valuable new additions which will assist them in managing this complex area, including the addition of an Adult Cyber-abuse Scheme and the Abhorrent Violent Material Blocking Scheme.

While these updates go some way towards protecting Australians online, there are further risks that are less clearly defined but are no less dangerous, including disinformation and its linkages to online harms.

Our submission proposes that we use this opportunity in updating the Bill to develop a boarder idea of what constitutes “online harm”.

To this end we have developed a device called the Online Safety Spectrum which attempts to map out the escalating range of online content and activity from safe to most harmful.



A spectrum more accurately covers the breadth of online harm affecting Australians today. We recommend the eSafety Commission and policy makers consider classification which more accurately reflects the reality of online harms.

We also recommend that the draft Online Safety Bill place greater emphasis on responsibility of the online platforms to take action, and for the eSafety Commission to

make full use its existing capabilities, including requesting wider and more transparent reporting from online platforms, and developing a wider range of industry codes and standards.

Introduction

The Australia Institute's Centre for Responsible Technology welcomes the opportunity to make a submission to consultations on a draft Bill for a new Online Safety Act.

The Centre for Responsible Technology was established to ensure sensible regulation leads to a safe, fair and healthy Internet. We share the eSafety Commission's core ambitions in improving and promoting online safety for Australians.

The draft Bill's enhancement and extension of the eSafety Commission's remit acknowledges the increasing need and demand for better online safety for Australians, in an environment of growing online risks.

While the draft Bill makes improvements to the current Online Safety Act and includes broader responsibility and powers for the eSafety Commission, we believe these updates do not go far enough in addressing the increased online harms we face today.

We are particularly concerned with the increased damage which disinformation causes, which directly and indirectly contributes to online abuses.

The updates to the Online Safety Act should reflect the many issues in online safety and encompass a broader range of dangers than what is currently being proposed.

The Spectrum of Online Safety

The eSafety Commission has remit over online safety and has powers to take action over online content which poses "serious harm" to Australians.

"Serious harm" in the draft Bill is defined as:

- serious physical harm or serious harm to a person's mental health, whether temporary or permanent
- serious harm to a person's mental health includes a) serious psychological harm; and b) serious distress[1]

The addition of new areas to the Online Safety Bill recognises the increased instances of online risks and has included "Adult cyber safety" and "Abhorrent, violent behaviour".[2] These additions are welcome.

Over the last couple of years, there have been concerning increases in online safety issues directly engaged with by the eSafety Commission, including:

- a 340% spike in online issues during the beginning of the 2020 COVID-19 lockdowns, including a 21% increase in cyberbullying, 48% increase in cyber abuse, and 86% increase in image-based abuse[3]
- 14% of Australians estimated to have been the target of online hate speech[4]
- Volumetric attacks, which are coordinated online attack campaigns, usually conducted by right-wing extremists and conspiracy theorists towards public figures promoting public health or community messages[5]

The eSafety Commission has successfully addressed many online harms over the last year period 2019/2020, including:

- 14,573 reports about potentially prohibited online content
- Identified 13,392 URLs hosting serious material including child sexual abuse material referred to law enforcement

[1] Australian House of Representatives (2020) *Draft Online Safety Bill 2020*

[2] Ibid.

[3] Medhora (2020), eSafety office records 340% spike in complaints as coronavirus impacts online behaviour, https://www.abc.net.au/triplej/programs/hack/complaints-esafety-increase-341-percent-because-coronavirus/12174654

[4] Koslowski & Lewis (2020), *What is hate speech?,* https://www.smh.com.au/national/what-is-hate-speech-20200202-p53wzy.html

[5] Calderwood (2020), *Extremists targeted Magda Szubankski with a trolling campaign after her Sharon Strzelecki COVID-19 ad*, https://www.abc.net.au/news/2020-10-22/right-wing-trolling-magda-szubanski-sharon-ad-covid-coronavirus/12800140

- 16 notices to services in relation to abhorrent violent material
- 690 complaints about serious cyberbullying[6]

While these are welcome developments the need for intervention is this space is growing. Cyber safety and risks are an everyday consideration for many Australians, who face a range of dangers online.

DISINFORMATION

A big part of this increased risk in online safety for Australians is disinformation.

Disinformation has become an insidious, wide-reaching risk for all Australians online. The major online platforms like Facebook and Google's Youtube are in an ongoing crisis in their inability to address disinformation in their systems.

Research conducted by the University of Canberra during the early days of the pandemic (April 2020) found that 23% of Australians encountered "a lot" of (dis)information and a further 36% encountered "some".[7] A report by the Centre for Responsible Technology in June 2020 found a coordinated network of conspiracy theorist clusters on social media disseminated disinformation about COVID-19.[8]

Disinformation is a serious issue, and can be the catalyst for more direct harm or abuse online. Recent examples demonstrate the link between disinformation and the defined online harms which the eSafety Commission is responsible for, namely; "serious physical and mental health harm" and "violent, abhorrent behaviour".

For example, disinformation on public health issues can lead to serious physical harm. During the COVID-19 period people have treated themselves with hoax cures and discredited medicines, as well as question the validity of vaccines.

During Victoria's recent COVID-19 outbreak, health experts warned that conspiracy theories were causing some Victorians to refuse testing (about 10,000 people).[9] Public figures with substantial online platforms, like Liberal member for Hughes Craig Kelly, and former celebrity chef Pete Evans, have been using their public platforms to spread conspiracy theories and anti-science falsehoods to Australians.[10] They both continue to have live

[6] ACMA and eSafety Office (2020), *ACMA and eSafety Office Annual Report 2019-20*

[7] Australian Communications and Media Authority (2020), *Misinformation an news quality on digital platforms in Australia position paper*

[8] Graham, Bruns, Zhu, Campbell (2020), *Like a virus, the coordinated spread of coronavirus disinformation*

[9] Razik (2020), *Health officials raise conspiracy concerns as 10,000 refuse coronavirus tests in Victoria*, https://www.sbs.com.au/news/health-officials-raise-conspiracy-concerns-as-10-000-refuse-coronavirus-tests-in-victoria

[10] Butler (2020), *Craig Kelly backs Pete Evans' right to spread conspiracy theories*: 'Ideas should be debated', https://thenewdaily.com.au/news/2021/02/01/craig-kelly-pete-evans-podcast/

presences on platforms like Facebook, Instagram and Twitter. Disinformation, whether spread by the general public or known figures contribute to ongoing dangers in public health.

Disinformation can lead to serious mental harm, like when Nationals MP Anne Webster became the victim of a malicious disinformation campaign against her and her husband. The MP so feared for her safety as a result of the falsehoods that she installed security cameras at her home. She describes the attack as leaving her "devastated" and "mortified".[11]

Disinformation can also lead directly towards violent and abhorrent behavior shown by the January 2021 US Capitol Hill riots. The riots were caused by a perfect storm of far-right militants and conspiracy theorists allowed to promote mob violence online, with several groups directly organised gatherings using the Facebook events feature. They were further incited by tweets from former President Donald Trump.[12]

The growing problem of disinformation cannot be cleanly sectioned off from other online safety risks, as disinformation can be one of the early catalysts for more direct action and harm later on.

The Centre for Responsible Technology therefore calls for the eSafety Commission and the Online Safety Bill to stretch its understanding of what constitutes online safety to include root causes of more serious harms.

Online safety should not just be acted on when the stronger, more direct symptoms have already manifested, like serious physical harm and violent behaviour, but must be treated at the cause as early as possible.

A useful device we developed is called the "Online Safety Spectrum", shown in Figure 1.

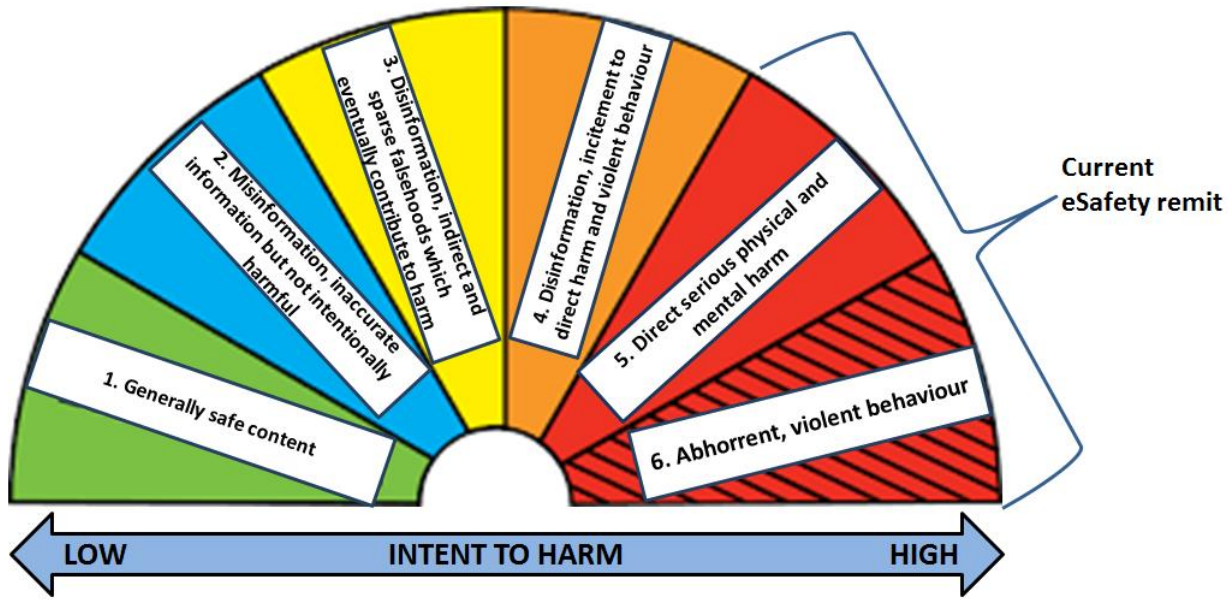
Inspired by other public safety alerts and categorisation, in this case the familiar Australian fire danger rating system,[13] a spectrum helps to broaden our understanding of what constitutes online safety and helps to demonstrate limitations of the Online Safety Act's current approach.

[11] McGowan (2020), *Australian MP installed security cameras out of 'genuine fear' after posts by conspiracy theorist, court hears,* https://www.theguardian.com/australia-news/2020/aug/31/australian-mp-installed-security-cameras-out-of-genuine-fear-after-attacks-by-conspiracy-theorist-court-hears

[12] Scott (2021), *Capitol Hill riot lays bare what's wrong with social media*, https://www.politico.eu/article/us-capitol-hill-riots-lay-bare-whats-wrong-social-media-donald-trump-facebook-twitter/

[13] NSW Rural Fire Service, *Fire Danger Ratings*, https://www.rfs.nsw.gov.au/plan-and-prepare/fire-danger-ratings

Figure 1. Online Safety Spectrum



In our Online Safety Spectrum we have broadly categorised online harms, with a range between one to six, one being generally safe, and six being the worst incarnations of abhorrent, violent behaviour. The range is judged overall by “intent to harm” from low on the left end of the scale to high on the right.

Category 1 encompasses generally safe content and activity online, which poses little to no risk.

Category 2 encompasses content that is broadly safe, but may have some inaccuracies, errors and falsehoods, but shared without malicious intent. This can be broadly defined as misinformation.

Category 3 includes mis/disinformation without a clear intent targeted towards a specific action, person or group, but can cast doubt on official or evidence-based information (e.g. this could be posting a series of general false claims, but not directed towards a specific person or group)

Category 4 includes disinformation with a clearer intent or action, such as organising a group to protest against public health measures (e.g. conspiracy theorists against proposed lockdown)

Categories 5 & 6 include the current remit of the Online Safety Act, where direct serious physical and mental harm, as well as abhorrent, violent behaviour has already occured. This could also be broadly defined as malinformation

Currently the eSafety Commission is empowered to act on serious, direct harms and malinformation, but not disinformation, which can be just as harmful.

The spectrum represents an initial attempt at navigating the complexity and range of online harms and should be further developed with subject matter experts, human rights groups and official bodies. Even in this early stage however, it can be useful as a tool to navigate and categorise the breadth of online harms that must be addressed. For example, the spectrum can be included in news articles dealing with online safety, during public health discussions, and as part of official public service initiatives addressing online safety. The spectrum can demonstrate how the current Australian Online Safety remit is only focused on one end of a wider spectrum.

There have been some attempts at categorising online safety risks, including with a focus on disinformation.

The Australian Communications and Media Authority (ACMA's) 'Misinformation and news quality position paper' lists a number of harms caused by (dis)information, from acute harms including:

- Health and safety which may directly threaten someone's health and personal safety
- Public panic and social disruption
- Electoral integrity
- Financial and economic scams and disruptions to business[14]

They also list more serious, and longer lasting chronic harms like:

- Negatively impacting trust in public institutions
- Eroding trust in professional sources of information like the media and scientific community
- Reduced community cohesion from increased polarisation and division[15]

The UK Government's 'Online Harms White Paper' acknowledges the nuances of online harm categorisation, noting that there are some harms with a clear definition (like terrorist content, hate crimes and incitement of violence), while other harms have a less clearer definition (including disinformation and trolling).[16] Nevertheless they have taken a more

[14] Australian Communications and Media Authority (2020), *Misinformation and news quality on digital platforms in Australia position paper*

[15] Ibid.

[16] UK Department for Digital, Culture, Media & Sport, (2019), *Online Harms White Paper*

holistic and wider-ranging view of “online safety” than what the current Australian Online Safety Bill has proposed.

There is room for the Online Safety Act to take a similarly broad view, and begin to acknowledge the more indirect and less defined (but no less dangerous) aspects of online harms.

Treating the symptoms and not the cause

The spectrum demonstrates that there is opportunity in treating the online harms at an earlier stage than what the eSafety Commission is currently set up for, and that consideration towards a 'prevention' rather than a 'cure' approach may be more effective and impactful ongoing.

The current process for the eSafety Commission is based on a 'complaints handling' or 'policing' approach. That is, victims are expected to report any issues after the fact, and the eSafety Commission is then empowered to address those direct complaints.

With online safety issues proliferating, this approach will increasingly become unsustainable, as a groundswell of online complaints could be bottlenecked by this process.

The eSafety Commissioner has acknowledged the challenge of attempting to address online harms in this way, likening the approach to:

> swatting individual bees in the midst of a swarm[17]

The Commissioner also rightly points out that:

> It is the platforms themselves who have the advanced technologies required to capture this form of online abuse at source.[18]

We recommend therefore, that policy makers use this update of the Online Safety Act, as an opportunity to develop a more future-proof approach of addressing online harms, developing devices like the Online Safety Spectrum, and placing more emphasis on the responsibility of the online platforms, giving consideration to tackling online harms at an earlier stage before the actual harm itself has already manifested.

[17] Biggs and Hunter (2020), *'Disturbing and harmful': eSafety commissioner calls on Facebook to stop 'volumetric' trolling*, https://www.smh.com.au/technology/disturbing-and-harmful-esafety-commissioner-calls-on-facebook-to-stop-volumetric-trolling-20200903-p55s33.html

[18] Ibid.

Reporting Determinations

The draft Bill will give the eSafety Commission the ability to request reporting from online platforms as part of a basic set of online safety expectations. In the draft Bill, this specifically entails asking online platforms to:

- Prepare periodic (and non-periodic) reports about the extent to which (online platforms like Youtube and Facebook) complied with the applicable basic online safety expectations
- Prepare periodic (and non-periodic) reports about the extent to which (online platforms like Youtube and Facebook) complied with one or more specific applicable basic online safety expectations[19]

It is not clear whether these reports will be publicly accessible in their entirety given potentially sensitive disclosures regarding specific cases and complaints.

As the use of online platforms grow and resulting risks of online harm grow, the eSafety Commission has the opportunity to be an authoritative source and voice in the education of Australians on the impact of platforms.

We recommend that these reporting determinations be used to their fullest capacity, and for the compiled information to be released to the public in an accessible manner, such as on the website of the eSafety Commission, as long as individual cases or victims are protected and no sensitive information regarding individuals are released.

It would be of public and community benefit if regular, up to date statistics on incidents, issues and occurrences of online safety harms across the largest online channels, like Youtube, Facebook, Twitter and Instagram are released, helping to educate all Australians on the live harms and dangers occurring on these platforms.

[19] Australian House of Representatives (2020) *Draft Online Safety Bill 2020*

Industry Codes & Standards

As part of the Online content scheme in the Online Safety Act, the eSafety Commission is empowered to:

- (call for) bodies and associations that represent sections of the online industry (to) develop industry codes
- Develop an industry standard
- Make determinations to regulate (online platforms)[20]

We recommend that these industry codes & standards powers be used to their fullest capacity.

Internet companies currently enjoy some of the lightest regulatory environments in the world, despite their massive influence in online safety, health and wellbeing and our overall information and public ecosystem.

The Australian codes around online safety need to match the reality of the risks of being online today.

Disinformation is of particularly concern, with Australia's only dedicated disinformation code being facilitated by tech industry lobby group DIGI.[21] The Centre for Responsible Technology in its submission to the voluntary code notes that self-regulation does not work and the proposed initiatives in the draft code were inadequate.[22] This code was due for a public update on February 2021, but as of the time of writing this report, no updates have been released.

The huge threat of disinformation cannot be properly addressed through a voluntary code, and the clear linkages between disinformation and online should demonstrate that disinformation could be part of stronger enforcement capabilities through the eSafety Commission.

At a minimum there needs to be much more than just one voluntary code to address this growing threat. Disinformation is a core part of the online safety landscape, and should be given more prominent attention and action, via enforceable industry codes and standards.

[20] Australian House of Representatives (2020) *Draft Online Safety Bill 2020*

[21] DIGI (2020), *Disinformation Code*, https://digi.org.au/disinformation-code/

[22] Guiao (2020), *Ensuring a strong and meaningful Code on Disinformation, Submission to the Australian Code of Practice on Disinformation*

Aside from this update to the Online Safety Act , and the voluntary code for disinformation, there are not enough industry codes and standards regulating the online landscape, and the eSafety Commission should exercise its full powers in facilitating these to be developed.

Recommendations

The Centre for Responsible Technology proposes the following recommendations for the updates to the Online Safety Bill:

- Develop a more sophisticated and broader definition of online safety, working with civic groups, independent policy bodies, experts and academia to use devices like the Online Safety Spectrum to create more holistic and effective online safety solutions.
- Place more emphasis on the responsibility of the online platforms in addressing online harms.
- Allow the public to view reports and determinations on compliance with standards, releasing regular reports on current and past incidents from online platforms.
- Make full use of their ability to develop industry codes and standards.
- Specifically address the growing issue of disinformation as part of the online safety landscape, more actively promote and develop the codes and actions around disinformation, including the voluntary code for disinformation currently being developed.

Conclusion

Online safety is a critical issue for Australians. Our pervasive use of network technology exposes us to regular online threats. The growing range of online harms is no longer just restricted to malinformation, but also the more widely known disinformation.

The current remit of the eSafety Commission and the existing proposed updates to the Online Safety Bill still only consider the extreme end of the online safety harm scale, and misses less extreme but still dangerous forms, like disinformation.

Our submission recommends that we stretch our understanding and definition of what constitutes online harms to better reflect the current reality, develop devices like the Online Safety Spectrum, place stronger emphasis on the responsibilities of online platforms, and ensure the eSafety Commission makes full use of its capabilities by requesting more robust reporting from online platforms and developing a wider range of industry codes and standards.