Australian Government | eSafetyCommissioner

Submission

Parliamentary Joint Committee on Law Enforcement Inquiry: Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019

15 October 2021

Contents

The eSafety Commissioner

The eSafety Commissioner (eSafety) is Australia's national independent regulator for online safety. Our core objective is to minimise harm to Australians online.

eSafety is the first government agency in the world dedicated specifically to online safety. We lead, coordinate, educate and advise on online safety issues and aim to empower all Australians to have safer, more positive online experiences.

When eSafety was formed in July 2015 (as the Children's eSafety Commissioner), one of the agency's main functions was administering a new regulatory scheme in relation to serious child cyberbullying. eSafety also assumed responsibility for the Online Content Scheme set out in Schedules 5 and 7 to the *Broadcasting Services Act 1992* (Cth) (**the BSA**), and previously administered by the Australian Communications and Media Authority.

Since then, eSafety's functions have broadened to include administration of a civil penalties regime in relation to image-based abuse ('IBA', sometimes referred to as 'revenge porn'), the power to issue notices to content and hosting services about abhorrent violent material, and a function related to blocking websites providing access to certain terrorist content during an online crisis event.

Beyond the protections built into our authorising legislation to provide takedown of harmful content and deliver compassionate citizen service, prevention through awareness and education and initiatives to promote proactive and systemic change are fundamental elements to our successful regulatory model.

In drafting this submission, we have had regard to items (a), (c) and (e) of the Inquiry's terms of reference, along with several related matters.

eSafety and the Christchurch attack

The tragic March 2019 livestreamed terror attack in Christchurch produced a range of responses from the Australian Government. The responses summarised below should be understood as components of an overall policy framework to limit the ability of terrorists and violent extremists to exploit the internet for propaganda purposes. eSafety hopes that, by providing an overview of the context for the use of abhorrent violent material (AVM) notices, we will inform the Committee's understanding of the full suite of operational considerations that attach to their use.

Shortly after the attack, the eSafety Commissioner began engaging with major Australian internet service providers (ISPs) in relation to the availability online of both the attack footage, and the attacker's manifesto. This included the creation of interim arrangements to facilitate the voluntary blocking by ISPs of a number of domains providing access to the material. The eSafety Commissioner also joined colleagues across Government at a 26 March 2019 Summit in Brisbane chaired by the Prime Minister that included representatives of the major digital platforms and ISPs to discuss the attack and industry responses.

Shortly after the Summit, the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (**AVM Act**) received Royal Assent.[1] Without extracting the relevant provisions for the

[1] The Act commenced 6 April 2019.

purpose of this submission, the AVM Act created a new power for the eSafety Commissioner to issue a written notice to a provider of a content service or hosting service notifying them that AVM is accessible from or hosted on their service. More about the AVM notice scheme is provided below.

One outcome of the Summit was the creation of a Taskforce to Combat Terrorist and Extreme Violent Material Online, jointly chaired by the Department of Prime Minister and Cabinet, and the Department of Communications and the Arts (as it was then known). Several recommendations of the Taskforce published in its June 2019 Report concerned eSafety. Relevantly, they included:

- **Recommendations 5.1 and 5.2 –** that the eSafety Commissioner should consider the temporary use of subsection 581(2A) of the *Telecommunications Act 1997* (Cth) (**Telco Act**) to
    - direct ISPs to continue blocking domains hosting the Christchurch attack footage and attacker's manifesto and
    - effect blocking of domains during an online crisis event.
- **Recommendation 6.2 –** that the eSafety Commissioner undertake the initial assessment of any content flagged in response to an online crisis event.

On 6 September 2019, the eSafety Commissioner issued the *Telecommunications (Protecting Australians from Terrorist or Violent Criminal Material) Direction (No. 1) 2019* (**the Direction**). The Direction formalised blocking action voluntarily taken by ISPs against websites providing access to the Christchurch material. This was a temporary, six-month direction that was put in place as an interim measure to stem the viral spread of the material and prevent it from causing harm to Australians, while government considered longer-term options for addressing misuse of online platforms by perpetrators of terrorism and violent extremism.

In December 2019, in conjunction with government and industry and in line with the Taskforce Report, eSafety finalised a protocol governing ISP blocking in an online crisis event (**the Protocol**). The Protocol sets out when the eSafety Commissioner will use these powers in relation to potential future 'online crisis events', defined as incidents involving terrorist or extreme violent material being shared widely online in a manner likely to cause significant harm to the Australian community. The Taskforce Report emphasised that such events would require a rapid, coordinated and decisive response by industry and relevant government agencies to contain the viral spread of the material.

The Protocol establishes detailed criteria, high thresholds and checks and balances to ensure the eSafety Commissioner's powers are used only in very limited and very serious circumstances. Any blocking direction made under the Protocol would only be in place for a limited time, to be determined on a case-by-case basis. Following the initial blocking period, the eSafety Commissioner could take further action to address the relevant material, in consultation with the ISPs and affected websites.

Abhorrent violent material: eSafety's role

eSafety currently exercises powers to investigate complaints from Australians about material that is prohibited or potentially prohibited under the Online Content Scheme.[2] In financial year 2020 – 2021 eSafety received almost 25,000 complaints about online content. About 65% of all

[2] Currently set out under Schedules 5 and 7 to the *Broadcasting Services Act 1992* (Cth), and due to be modernised through commencement of the *Online Safety Act 2021* (Cth) in January 2022.

complaints concerned child sexual exploitation material (CSEM), including content depicting the rape and torture of children. Other types of illegal and seriously harmful material represented in complaint data included material that promotes, instructs, or incites in matters of crime, extreme violent material, and material advocating the doing of a terrorist act. Under the current scheme, the eSafety Commissioner has the power to issue a takedown notice in relation to prohibited material hosted in Australia. Information about the updated Online Content Scheme, enabled by the *Online Safety Act 2021* (Cth) (**the OSA**), is provided below.

Under the AVM Act, the eSafety Commissioner may issue a notice to a content service (e.g., a social media service or website) or a hosting service which *puts the service on notice* that they are providing access to or hosting AVM. It should be noted that AVM is defined very narrowly as audio, visual, or audio-visual material produced by a perpetrator or their accomplices, which a reasonable person would regard as offensive, and depicts the perpetrator(s) engaging in:

- a terrorist act involving serious physical harm or death
- murder or attempted murder
- rape
- torture, or
- kidnapping involving violence.

It should be emphasised that the scheme is not a takedown regime. That is, an AVM notice is not enforceable by the eSafety Commissioner in the fashion of a regulatory takedown notice. Instead, the AVM notice serves to establish certain evidentiary requirements necessary for a criminal prosecution against a provider that it was reckless as to the presence of AVM on its service.

Management of AVM notices

Since the introduction of the AVM amendment, eSafety has undertaken 2,120 regulatory investigations under the Online Content Scheme into material satisfying the definition of AVM. All of this material has been hosted overseas, and has overwhelmingly (more than 98%) depicted penetrative child sexual abuse or torture. These items form a component of the 32,000 regulatory investigations that the eSafety Commissioner has concluded into various types of CSEM since passage of the AVM Act. In all but a handful of matters, we notify CSEM to member hotlines of the Internet Association of Internet Hotlines (INHOPE),[3] for takedown in the host jurisdiction. Almost 75% of CSEM URLs are removed by INHOPE members within three days of notification, and in 2020 INHOPE exchanged one million CSEM URLs.

The efficiency and rapidity of CSEM removal within the INHOPE network represents a useful alternative to other notification options (including AVM notices). eSafety's participation in the network is enabled via a memorandum of understanding with the Australian Centre to Counter Child Exploitation (ACCCE).

From public complaints about illegal and harmful online content, eSafety has identified a relatively small number of items that are amenable to AVM notices. Since April 2019, eSafety has issued 24 notices to a variety of content and hosting services in relation to 15 items of content

[3] INHOPE is a global network of 46 hotlines dedicated to combatting CSEM. The network includes the National Centre for Missing and Exploited Children in the United States, and the UK Internet Watch Foundation.

depicting fatal terrorist violence, murder and torture. Content has been removed by a service in 87% (13 out of 15) of matters where AVM notices have been issued.

eSafety has discretion to determine whether an AVM notice should be issued to the content service, the hosting service, or both. This discretion allows eSafety to serve notices that have the best chance of reaching the intended service provider and reflects that there are multiple parties – that is content services, upstream hosting providers and their downstream clients or subsidiary – involved in making content available online.

Not all items of content that satisfy the definition of AVM will necessarily be acted on via an AVM notice, and eSafety exercises discretion based on criteria such as:

- whether the material is capable of being shared widely
- whether the material has the potential to incite violence or cause harm to victims, their families or the broader community
- the recency of material
- whether the material appears to serve an extremist agenda or propaganda purpose and
- the nature of the relevant service(s), including whether there is an identifiable means of giving a person (legal or natural) an AVM notice.

As part of our standard operating procedures, eSafety advises the Australian Federal Police (AFP) when intending to issue an AVM notice, and again once a notice has been issued.

The majority of items eSafety has notified to content services and hosting services relate to versions of the Christchurch attack footage.[4] Examples of other content types include:

- images showing the immediate aftermath of the murder of a 17-year-old female
- video depicting the torture by flaying of an adult male
- versions of a video depicting the murder of two Scandinavian tourists in Morocco and
- videos depicting the murder of two people by a terrorist in Halle, Germany.

AVM response example: eSafety's response to the Halle, Germany, attack

On 9 October 2019, a single attacker used the gaming platform Twitch to livestream an unsuccessful military-type assault on a synagogue in Halle, Germany. Prior to this terror attack taking place, the alleged perpetrator posted a manifesto online.

As soon as eSafety became aware of the incident on 9 October Sydney time, we began engaging with executives at Twitch and its owner, Amazon, along with the other major digital platforms.

The eSafety Commissioner was satisfied that the threshold for an online crisis event had not been met and therefore ISP blocking was not considered. Factors taken into account included:

(a) the distribution of the attack footage was relatively limited compared, for example, to the virality of the Christchurch attacks

(b) the material was not being successfully uploaded on Facebook, Twitter or YouTube

(c) given the spread of the content was contained, the risk of significant harm to the Australian community was low

[4] It should be noted that the Christchurch attacker's manifesto would be not considered AVM given that it is not audio, visual or audio-visual material depicting relevant perpetrator-produced abhorrent violent conduct.

(d) engagement with global partners, such as the Global Internet Forum to Counter Terrorism (GIFCT), which made a similar determination in relation to the content and did not activate their Content Incident Protocol.

While there were no public reports made to eSafety about the livestream video, eSafety investigators identified related material that could be accessed by Australians on three websites. Owing to the seriousness and immediacy of the matter, eSafety issued notices to each of the content services and related hosting services. All items of content were removed.

AVM notices and other elements of eSafety's remit

eSafety views the AVM notice scheme as complementary to its existing (and, through the OSA, future) powers. As noted above, the AVM notice power went some way to filling a regulatory gap, where eSafety was not empowered under the BSA to act against overseas-hosted prohibited violent terrorist or extremist material.

While the AVM notice power is not a takedown power, close to 90% of material has been removed following the receipt of a notice. eSafety has not collected reasons from all service providers who have elected to remove material. However, we do know that the link between the notice scheme and criminal penalties has clearly incentivised removal. In several matters, we have been told by services (or have learned through statements made elsewhere) that the risk of criminal prosecution was the only reason for their removal of material.

eSafety regards the scheme as an effective deterrent, limiting the viral spread of violent terrorist propaganda online. The fact that notices have been relatively few in number should, we believe, be seen as further evidence for this deterrent effect.

Online Safety Act 2021

A major reform to the regulation of online harms will commence in January 2022 through the OSA. The OSA is intended to create a modern, fit for purpose regulatory framework that builds on the existing legislative schemes for online safety. Relevantly the OSA:

- strengthens the existing Online Content Scheme by expanding the number of services relevant to its operation, and providing the eSafety Commissioner the power to issue removal notices against 'class 1' content (which includes pro-terror content and AVM) wherever that content is hosted globally
- creates the power to issue a blocking notice to an ISP in relation to material provided on a web domain that is relevant to an online crisis event
- creates new powers for the eSafety Commissioner to direct online app stores and providers of online search services to remove apps and delete links that allow access to that material where one or more class 1 removal notices have been ignored
- introduces a set of Basic Online Safety Expectations through a ministerial legislative instrument that will allow the eSafety Commissioner to require reporting on how services are keeping their users safe, which could include how they are preventing their platforms from being used as a vehicle for pro-terror content and AVM, and
- provides for the creation of one or more industry codes or standards to promote the adoption of responsible industry processes and procedures for dealing with online content issues, including pro-terror content and AVM.

In addition, the OSA creates a world-first scheme to address seriously harmful adult cyber abuse, enhanced child cyberbullying and image-based abuse schemes, and improved information-gathering powers. eSafety has produced a fact sheet on the OSA, available here.

With respect to the interaction between AVM notices and the OSA, there may be circumstances where one regulatory approach is preferred over the other. Even so, there is nothing to preclude the Commissioner from issuing class 1 removal notices and AVM notices concurrently. Where dealing with a content service located outside Australia, the threat of possible prosecution may hold more weight than a civil penalty order.

Blocking notices under the OSA

Part 8 of the OSA provides the Commissioner with powers to request or require ISPs to block access to material that promotes, depicts, incites or instructs in abhorrent violent conduct. The intent of this power is to protect the Australian community by preventing the rapid and widespread distribution of terrorist and extreme violent material during an online crisis event.

Failure to comply with a blocking notice under the Act can lead to civil penalties.

Before issuing a blocking notice, the Commissioner must be satisfied that availability of the material online is likely to cause significant harm to the Australian community. In considering whether this is the case, the Commissioner is to have regard to the nature of the material (for example, whether it is livestreamed material, particularly high impact material such as a beheading), and the potential for the material to go viral on the internet (i.e. the numbers of end-users who are likely to access the material).

There are exemptions for material that relates to news reports that are in the public interest, ensuring that the power is sufficiently targeted, and is crafted in a way that is reasonable, proportionate and necessary to achieve the policy objective of online safety for Australians.

In deciding whether to exercise the power, the Commissioner must have regard to whether any alternative options could be used to minimise the likelihood that the availability of the material online could cause significant harm to the Australian community.

The ISP blocking power is intended to enhance and complement eSafety's other powers, including those relating to removal of class 1 content and notification of AVM. Under the OSA, eSafety will be able to choose from a range of options in an online crisis event, tailoring our response to the circumstances.

Class 1 content

Under the OSA, the eSafety Commissioner will have the power to issue removal notices against class 1 content, even though the content may not be hosted in Australia. Material will be considered class 1 material when it has been or is likely to be classified RC (Refused Classification) by the Classification Board. As a consequence, class 1 material includes child sexual exploitation material, material that provides detailed instruction or promotion of in matters of crime or violence, material that advocates the doing of a terrorist act, and material that depicts certain types of violence (such as very detailed or very high impact depictions of cruelty and real violence).

eSafety will be working with the Communications Alliance and ISPs in the coming months to update the Protocol to reflect the new legislative provisions and ensure that arrangements are in place to activate these provisions quickly if necessary.

Basic Online Safety Expectations

The Act establishes core Basic Online Safety Expectations (expectations) for online services. The expectations will help ensure that these services are safe for Australians to use and provide greater transparency around their safety features, policies and practices.

The expectations will be set out in a determination made by the Minister for Communications, Urban Infrastructure, Cities and the Arts (Minister). eSafety will have the power to require services to report on their compliance with any or all of these expectations, and to publish statements about their compliance.

The Department of Infrastructure, Transport, Regional Development and Communications is currently undertaking public consultation on a draft determination. The draft determination includes an expectation that services will take reasonable steps to minimise the extent to which material that promotes, incites, instructs in or depicts abhorrent violent conduct is provided. It also includes an expectation that services will have clear and readily identifiable mechanisms that enable end-users to report and make complaints about such material.

Industry Codes

Additionally, the OSA requires the development of industry codes that will seek to limit exposure to the most harmful type of online content, including pro-terrorist material and extreme violence.

The codes will apply to the participants of eight key sections of the online industry, including providers of social media, messaging, search engine and app distribution services, as well as internet and hosting service providers, manufacturers and suppliers of equipment used to access online services and those that install and maintain the equipment.

The eSafety Commissioner must make reasonable efforts to ensure that a code for each section of the online industry is registered within six months of commencement of the Act.

eSafety has released a position paper to assist the online industry to develop codes.

The paper sets out 11 policy positions regarding the substance, design, development and administration of industry codes, as well as eSafety's preferred outcomes-based model for the codes.

eSafety has engaged with a number of industry bodies and associations to date and will continue to work closely with industry to ensure codes are registered within the legislative timeframe.

Safety by Design

Finally, eSafety has spearheaded the global roll-out of the Safety by Design initiative. Safety by Design focuses on the ways technology companies can minimise online threats to users – especially younger users – by anticipating, detecting and eliminating online harms before they occur. Embedding safety into online products and services as core features from the very outset of product design is at the heart of the Safety by Design ethos.

Key to the initiative is a framework built around principles covering platform responsibility, user empowerment, and transparency and accountability. The principles have now been translated

into a set of comprehensive tools allowing companies – from start-ups to established enterprises – to evaluate the current safety of their systems, processes and practices. The tools were developed with and for industry, highlighting industry best practice in innovations for safety.

Through Safety by Design, eSafety is seeking to lift the safety standards and practices of the technology industry to ensure greater protection of users and to minimise future threats. Safety by Design is intended to shift responsibility back to the platforms for safeguarding their users and engineering out misuse before harm occurs, rather than retrofitting fixes once the damage has been done.

Observations: operation of the AVM Act

Industry response

As noted above, eSafety has seen a high degree of success with AVM notices. It might be observed that none of the notices issued to date concern a major digital platform. Our experience of managing AVM notices is that the platforms are responsive to informal approaches by eSafety when we identify AVM or links to AVM on their services. For example, recently eSafety responded to a notification that the Christchurch attack video was linked to from a social media service. We contacted the social media service to alert them to the link, and it was removed swiftly and without the need to issue a formal notice. The link was also banked to prevent further attempts to share the material via the platform. To date, eSafety's experience working with platforms on this issue has been a constructive one.

eSafety's functions and investigation powers

This submission refers elsewhere to the fact that an AVM notice is not a takedown notice. In other words, eSafety does not have the same regulatory powers to enforce removal as are available in relation to Australian-hosted content under the BSA and class 1 content under the OSA (which applies globally). The notice issued by eSafety serves to help prove that a provider was aware of AVM on its service, and was reckless as to that fact, with recklessness the relevant fault element under the offence provision at section 474.34 of the *Criminal Code Act 1995* (Cth).

While this provides a useful link to a criminal enforcement regime, the notice power is not a power that is established as one of the eSafety Commissioner's statutory functions through the *Enhancing Online Safety Act 2015* (Cth), nor is it a regulatory power per se which comes with the ability to exercise information-gathering or investigation powers.

Under the BSA, an investigation into online content can be undertaken to determine whether, for example, Australian end-users can access prohibited material provided by a content service, or whether a hosting service is hosting prohibited material. However, once it becomes apparent that there is no Australian connection to the content (as required under the BSA for a takedown notice) – meaning no takedown notice is possible – the exercise of compulsive powers merely to facilitate service of an AVM notice raises potential legal questions.

Under the OSA, the overlap between class 1 content and AVM is stronger. Most forms of material depicting abhorrent violent conduct are likely to be Refused Classification. However, this does not apply to all forms. Consider the example of a video showing a violent incident where a person is forcibly removed from a vehicle, punched, bound and then bundled into a van. On its face it appears that this may be a kidnapping, but the level of violence is likely to fail the class 1 threshold test. However, if eSafety knew that the video was produced by an accomplice to a serious crime, such as murder, rape, kidnapping or an act of terrorism, the material would be AVM. In such a case, where eSafety is considering an AVM notice (but not a class 1 notice) eSafety would not be able to exercise formal investigative powers.

The lack of compulsive information-gathering and investigative powers can result in difficulties. These difficulties are especially evident in cases where a hosting service notice is contemplated that involves networks operated by large providers of internet infrastructure. In those cases, the network IP address returned from searches of public databases (such as the WHOIS database, assuming the relevant entry is accurate) may be the responsibility of a corporate entity which lies 'downstream' from the entity to which it is assigned. Given the lack of specific investigative *powers* which attach to the AVM notice power, eSafety cannot currently seek to compel documents or other information from the upstream provider to identify their downstream client or subsidiary. Challenges thus arise in relation to ensuring the accurate drafting of notices, which must correctly specify the corporate entity against which evidence of recklessness might lie. This may not always be possible when company details are not visible from public records.

eSafety's AVM notices and differing standards of proof

eSafety has no role in any criminal investigation or prosecution of AVM notices issued under 474.35 and 474.36 of the Criminal Code. The notices issued by eSafety give rise to presumptions in relation to some, but not all, of the elements of the offences against subsections 474.34(1) and 474.34(5) of the Criminal Code.

Such a notice may be issued when the eSafety Commissioner is satisfied on reasonable grounds that at the time the notice was issued:

> (a) the specified content service could be used to access the specified material; and
>
> (b) the specified material was abhorrent violent material.

However, these presumptions are rebuttable if a defendant adduces or points to evidence that suggests a reasonable possibility that, at the time the notice was issued, the person was not reckless as to whether the specified material was abhorrent violent material.

Once the presumption is rebutted proof will revert to the criminal standard of beyond reasonable doubt.

There may be a tension between the standard required to issue a notice and the higher standard of proof required in any subsequent criminal prosecution, particularly relating to:

- Whether the notice was received; and
- The substance of the notice itself.

There may be utility in considering whether the change in threshold could have an impact on a prosecution.

Additionally, as discussed, eSafety has no compulsive investigation powers that relate directly to the investigation of an AVM notice unless the material also meets the definition of class 1 material under the OSA. As a result, eSafety may need to rely on assistance from the AFP to investigate more fully. eSafety and the AFP continue to engage productively on a set of a procedures to facilitate this, however, this review presents an opportunity to consider whether there would be benefit in having additional guidance on this point.