Microsoft Submission to the Parliamentary Joint Committee on Law Enforcement

Inquiry into the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*

October 2021

Introduction

Microsoft welcomes the opportunity to present this submission to the Parliamentary Joint Committee on Law Enforcement (**PJCLE**) in relation to its ongoing inquiry (**Inquiry**) into the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (Cth) (**AVM Act**).

We recognise the important role that technology companies play in combatting the spread of abhorrent violent material online (AVM). In addition to our responsibilities under Australian law, Microsoft seeks to address the spread of this content through a range of voluntary measures, including the use of tools and technology and working in collaboration with industry, civil society, and governments. Examples include our ongoing support for the *Christchurch Call to Action to eliminate terrorist and violent extremist content online* and the Global Internet Forum to Counter Terrorism (GIFCT), as well as our participation in the Australian Taskforce to combat terrorist and extreme violent material online.

We note that the rapid timeline in which the AVM Act was legislated prevented important prior consultation with relevant industry, civil society, and other affected groups. While Microsoft appreciates that the role of technology in the March 2019 Christchurch attack called for swift action, we consider that consultative and considered approaches to law reform result in more sustainable and effective legislation. For this reason, Microsoft welcomes the Inquiry into the AVM Act. Whilst we acknowledge this review is required under the legislation, we encourage the Committee to undertake a rigorous process in which the detailed consultation processes and mechanisms that typically apply to legislation with significant consequences are applied this time around.

Microsoft also notes that such rapid, non-consultative approaches to law reform can have a negative impact on investment and innovation in Australia. Not only did the rapid passage of the AVM Act create business uncertainty locally, but it also contributed to the idea that the regulatory landscape in Australia is unpredictable, particularly in the technology space. This is not consistent with the Government's stated ambition to be a leading digital economy. While the AVM Act may be 'world-first', this does not necessarily translate into effective outcomes for users in practice, proportionate expectations on industry, nor achievement of Government's legislative goals.

In this submission, we seek to explain our primary position, which is that the regulatory lacuna that prompted the introduction of the AVM Act has, since the time of its introduction, been addressed through various legislative, technical and international cooperative efforts. On this basis, we suggest that the AVM Act could be repealed without materially weakening the Commonwealth Government's stance on, or tools to deal with, the challenge of terrorist and violent extremist online harms.

In addition, and acknowledging the Terms of Reference of this Inquiry, we also outline several issues and recommendations relating to the AVM Act. While our primary position is that the need for the AVM Act

has been obviated since commencement, if it is to continue in force, there are several important areas where reform may enable the AVM Act to better serve its aims, as well as to increase certainty for both online service providers and the individuals using those services.

1. The AVM Act and its relevance today

Following the March 2019 terrorist attack in Christchurch, New Zealand, the Commonwealth Government led the rapid passage of an AVM bill through both houses of Parliament, with the AVM Act coming into effect on 6 April 2019. The AVM Act amended the *Criminal Code Act 1995* (Cth) (**Criminal Code**) to include new obligations and offences in relation to AVM, as well as enforcement powers for the eSafety Commissioner.

Microsoft fully appreciates that the AVM Act was passed in an effort to address a specific policy problem highlighted by the Christchurch attack - namely, how to prevent and swiftly stop the livestreaming of a violent real-world event, and subsequently, deliberate attempts to make content from that attack go viral. Two years later, the regulatory landscape has evolved, and a number of other structures and processes have been developed in response to the challenge highlighted by that attack. Accordingly, there is an opportunity to look at the AVM Act afresh and consider whether it remains an appropriate and proportionate response to material depicting abhorrent violent conduct (AVC) and AVM. We understand the powers in the Act have been very seldom used since its passage: as at 24 March 2020, the eSafety Commissioner had issued only 18 notices.[1]

We outline three measures below that did not exist at the time that the AVM Act was drafted, and that Microsoft believe substantially address the risks targeted in the AVM Act. While we do not deny that the potential remains for some technologies to be misused by motivated terrorist or violent extremist actors, the regulatory gap that existed in 2019 has since been filled. Moreover, as these measures were produced with valuable community consultation and, in some cases, have had their effectiveness proven in practice, we submit that the Inquiry ought to consider whether the AVM Act (and the provisions it inserts into the Criminal Code) remains a necessary and proportionate response. We believe it does not.

Measures that fill a previous regulatory gap:

a) Online Safety Act

Over the past two years, Microsoft has been an active participant in ongoing consultations regarding the *Online Safety Act 2021* (Cth) (**OSA**), which will commence on 23 January 2022. We recognise that the OSA will play an important role in addressing various forms of online harm, including terrorist content and content depicting extreme crime and violence. Importantly, the OSA includes a "modernised online content scheme", replacing that which was previously contained in Schedules 5 and 7 of the *Broadcasting Services Act 1992* (Cth) (**BSA**).

For instance, Part 8 of the OSA includes new blocking powers that can be directed at internet service providers to block certain domains that contain terrorist and extreme violent material. These provisions precisely target 'online crisis events' like the Christchurch attack, can be effective in preventing harm to a large number of end users, and are drafted in more proportionate and workable terms than the AVM Act. The type of material captured by these new powers is also wider than the AVM Act provides for, meaning that it is capable of addressing harm associated with both 'AVM' as well as other types of terrorist and violent extremist material that may not reach the strict definition of AVM. The OSA was also carefully developed over a long period of time, with multiple staging points for public comment.

[1] https://www.esafety.gov.au/sites/default/files/2020-03/eSafety-AVM-factsheet.pdf

In addition, under the OSA, the eSafety Commissioner will have the ability to issue timed removal notices for Class 1 material (the definition of which would encompass AVM) regardless of whether that material is hosted in Australia or not. The Explanatory Memorandum to the OSA even acknowledges that this approach is consistent with the extraterritorial powers provided by the AVM Act. This aligns the treatment of AVM with other serious, illegal content such as child sexual abuse material. The OSA also introduces a range of penalties for non-compliance.

The OSA's Basic Online Safety Expectations (**BOSE**) Determination will also create express obligations with regard to material that encompasses AVM, which will be complemented by the industry codes currently under development. The industry codes will also require providers to take measures to address class 1 content, with the measures proportionate to the risk faced by a service.

Through the measures outlined above, as well as the increased scope and powers it offers to the eSafety Commissioner, the OSA represents a dramatic step-change in Australia's regulatory approach to online harms. This being the case, the regulatory landscape at the time the AVM Act was introduced has substantially and substantively shifted, warranting serious consideration as to its continued utility and appropriateness. Retaining a separate AVM regime also risks creating inconsistencies.

The "Report of the Statutory Review of the EOSA and Schedules 5 and 7 of the BSA" in 2018 (otherwise known as the **Briggs Review**) recommended the consolidation of the eSafety Commissioner's powers, which had previously been spread across the *Enhancing Online Safety Act 2015 (Cth)* (**EOSA**) and BSA, and later the AVM Act. The Briggs Review also envisaged the benefit to both industry and community outcomes that a single piece of legislation could provide.

b) GIFCT and the Content Incident Protocol

Acknowledging that the Christchurch attack warranted new attention toward perpetrator-produced online content, the GIFCT mobilised to develop the Content Incident Protocol (**CIP**), consistent with the commitments of the Christchurch Call to Action. The CIP is the process by which GIFCT member companies quickly become aware of, assess, and address potential content circulating online resulting from an offline terrorist or violent extremist event.

Once the CIP is activated, hashes of the related content are shared in the GIFCT's hash database to support member platforms to be able to detect that content and respond swiftly. The CIP is activated where:

(a) a real-world terrorist, violent extremist or mass violence event;

(b) has been recorded or broadcast via livestream;

(c) depicting murder or attempted murder; and

(d) is being distributed on the GIFCT member platforms or so broadly online that such distribution appears inevitable.

The GIFCT is also continuing to build out its Incident Response Framework beyond video and live-streamed content, to help facilitate information-sharing amongst member companies and to establish procedures to respond to other forms of online materials shared by a perpetrator or accomplice as well. The GIFCT has responded to over 140 incidents since 2019, with member companies sharing information and situational awareness to understand if an attack has a particular online dimension. Coordination between GIFCT member companies (which include some of the world's largest communications platforms) enables a practical, global response.

Moreover, work to enhance crisis response procedures also continues in the context of the GIFCT Crisis Response Working Group, which was co-facilitated by Microsoft in 2020. We have been pleased to work in a multistakeholder group (including with governments) on practical steps to enhance crisis response procedures.

c) *Australia's new Online Content Incident Arrangement*

In March 2019, the Commonwealth Government established the Taskforce to Combat Terrorist and Extreme Violent Material Online (the **Taskforce**) which was to provide advice on practical, tangible and effective measures to combat the upload and dissemination of terrorist and violent extremist material. Flowing on from the report delivered by the Taskforce in June 2019, the Department of Home Affairs established the Online Content Incident Arrangement (**OCIA**) alongside the eSafety Commissioner, government agencies, and industry (including Microsoft).

The OCIA helps provide a clear framework for cooperation between government and industry to respond to AVM online following a crisis event and has been tested through a virtual tabletop exercise. While all incident response processes are subject to continuous review and improvement, the establishment of the OCIA helps demonstrate a new level of public-private coordination to respond to the circulation of the kind of AVM that prompted the passage of the Act.

2. Issues with the AVM Act and its present implementation

The issues that sit at the core of the AVM Act are deeply complex and warrant commensurate levels of consultation with industry and other stakeholders in the online space to ensure positive outcomes in the real world. Below are several issues we have identified over the time in which the Act has been in force.

Challenges in identifying AVM

The AVM Act provides very targeted definitions of abhorrent violent conduct (**AVC**) and the abhorrent violent material (**AVM**) that depicts it. Microsoft considers the specificity of these definitions to be appropriate in context and aligned with the purposes of the AVM Act. As was noted in the second reading speech for the AVM Act, "*it is important that this offence is limited to the worst types of material that can be shared online*".

However, the period since commencement of the AVM Act has highlighted how these legislative definitions must not just be precise, but also workable in real-world settings. This is particularly so given the increasingly demanding and sometimes conflicting regulatory obligations imposed on industry – not only in Australia, but worldwide.

Timeframes for the removal of AVM or notification of Australian-based AVC to the AFP are somewhat ambiguous and indeterminate, but in any case require "expedition". This fact, paired with the substantial liability exposure created by the AVM Act (including threat of criminal sanctions), places service providers into situations that incentivise a broad interpretation of AVM over the specific definition in the AVM Act itself.

In essence, while the definitions may be appropriate on paper, when they have to be considered in real-time alongside the obligations imposed by the AVM Act, critical nuance is undermined. Consider the following hypothetical scenario:

> ***Scenario 1****: A political protest overseas turns violent and a protester uploads footage of herself engaged in a physical altercation with a rival protester. Factually speaking, the video does not meet the statutory definition of AVM. Nonetheless, a concerned member of the public reports the*

video to the content service provider and flags it as terrorist content..

The service provider immediately and permanently takes down the video. The service provider explains that where a complaint matches certain AVM-related terms, it is immediately taken down and later assessed through human-review during business hours. Due to the poor quality of the video and lack of corroborating information, the employee who reviews the relevant video is unable to conclusively determine whether the content is AVM or not. However, to meet the 'expeditious' time requirement and avoid any risk of liability, the service provider has developed an internal policy that says human review must be complete within 15 minutes and any indeterminate material will be removed from the service.

Scenario 1 illustrates how targeted drafting can be displaced in practice due to the combination of high-pressure statutory obligations, resource-constraints (particularly for smaller or newer companies), and ambiguous visual media. The AVM Act as currently drafted creates a tendency to err on the side of removing content, rather than engaging with the intended legislative aims.

In addition to the difficulty in practically identifying AVM, we also note the following factors that can complicate the identification of AVC and AVM in real-world settings:

- In relation to AVC, the **geographic location** depicted in a video is not always evident from its contents or associated information (such as attached descriptions or public metadata). This may result in inappropriate reporting to the AFP under s 474.33 of the AVM Act, negatively impacting resource capacity of both the service provider and the AFP.

- The **identity of the producer** of a piece of content or any link to the perpetrator is not always evident from its content or associated information, complicating the satisfaction of s 474.31(1)(c).

- Material may visually appear to depict AVC despite actually being **fictional or artificial**, such as in the case of scripted entertainment media or realistic video-games.

- Other factors critical to the statutory definition of either AVC or AVM may be ambiguous and unable to be reliably verified with respect to a given piece of content.

The negative impact generated by these issues is not just limited to inconvenience or expense to industry. The AVM Act, unless properly tooled through consultation, is prone to implementation that may **restrict legitimate speech**, journalism and other lawful and important societal functions. For instance, footage of a shooting by a law enforcement officer may technically meet the definition of AVM but may be important for public discussion or in the context of a trial.

Indeed, the Commonwealth Government may be said to have foreseen the above issues and addressed them through the numerous defences (s 474.37) and prosecution safeguards (s 474.42). However, in practice, service providers are highly reluctant to take the risk of misidentifying material and risk becoming subject to severe financial penalties or imprisonment orders affecting local employees.

Recommendations regarding identifying AVM

If the AVM Act is retained, the issues identified above could be significantly ameliorated if the identification and classification of AVC and AVM was carried out by a delegated decision-maker, such as a judicial officer or appointed regulatory body. Such an approach would introduce transparency to a process that is currently highly discretionary, decentralised, and opaque. It would also improve the ability for industry-participants to efficiently deal with offending content, as additional time would not be

required to be spent by each participant making the complex assessments. Indeed, the OSA allows for this by giving the eSafety Commissioner the powers to order service providers to remove content within 24 hours. In general terms, we also consider it more appropriate for elected officials, independent courts or other parts of government to make decisions on the legality of specific content, rather than private companies. For instance, judicial decision-makers are the guarantors of due process for Australians, enabling a careful balancing of rights and the application of local law and human rights principles.

In the alternative, we believe that the challenges raised above could also be reduced by amending certain elements of the AVM Act. Practical expectations around takedown times could be made clearer and more proportionate, reducing the undue urgency that is undermining implementation efforts. The penalty regime should also be adjusted to better accommodate good-faith industry participants attempting to achieve compliance, while reserving the most severe penalties for repeat offenders and bad-faith actors.

Unclear takedown scope over time

The AVM Act originated in response to the Christchurch attack and the particular social harm of its online livestreaming and subsequent viral distribution. Since enactment of the AVM Act, we have noticed confusion amongst Government agencies, industry participants and public authorities alike in relation to how the AVM Act applies to relevant material over time. The fact that the severe consequences for providers in the AVM Act ensures the issue "gets the attention" of providers should not be considered a measure of success as it abrogates proportionality and due process, and if it fails to take into account unintended consequences. Microsoft takes seriously all its regulatory obligations, regardless of the penalty scheme.

AVM, as defined under the AVM Act, depicts a particular instance of AVC that has occurred in the real-world at a particular point-in-time. Once made available online, the unfortunate reality is that most AVM will stay in circulation or be able to be shared somewhere on the Internet on an ongoing and indefinite basis. Even small alterations to AVM can pose challenges to the detection tools used by responsible actors. There are also a range of other platforms that do not seek to moderate this content – and indeed, some that encourage its distribution. If retained, we see the AVM Act obligations as being most critical and sensible during or directly following an occurrence of AVC – i.e., a crisis event. As time elapses and an AVC event becomes historical, it is unclear how certain AVM Act obligations will continue to apply or whether it is indeed appropriate to continue to treat AVM the same way – especially given the new OSA regime mentioned earlier.

The notification obligations in s 474.33 of the AVM Act provide a logical caveat to this end, as subsection (3) notes that the obligation does not apply if the industry participant reasonably believes that details of the material are already known to the AFP.

The removal / ceasing to host obligations in s 474.34 are not similarly caveated. As such, a plain reading of the AVM Act suggests that this provision will apply to a piece of AVM on an ongoing and indefinite basis, regardless of how much time has elapsed since the occurrence of the related AVC or the reduced virality associated with that piece of AVM.

However, given that the fault element for AVM Act offences is recklessness, and that industry participants need not have actual knowledge of AVM to be in breach, Microsoft sees s 474.34 waning in its effectiveness, workability and proportionality over time in relation to a particular piece of AVM. As noted above, new tools will also soon be available to require the removal of such content.

This is further complicated by the fact that s 474.34 does not limit the jurisdictional origin of AVM. As

such, applicable industry participants are indefinitely exposed to serious liability if AVM from any place in the world, or point-in-time, is available on their service and accessible in Australia. And as noted earlier, malicious actors may alter the content to avoid detection. Consider the following hypothetical scenarios:

> ***Scenario 2**: An applicable service provider takes its AVM Act responsibilities seriously, and has implemented strict protocols to reduce the likelihood that known AVM is uploaded to their service. One day, a user uploads previously unpublished footage from an overseas conflict in 1980, depicting graphic AVC that was produced by the perpetrator. As the footage has never been published previously, it is not identified via the service provider's hash database. The material is accessed by a child in Australia, leading to a complaint that brought the material to the service provider's attention. The service provider has committed an offence.*
>
> ***Scenario 3**: An applicable service provider has previously removed AVM that first originated after a 2019 occurrence of AVC. It has subsequently implemented technical measures to prevent further upload of that AVM, which has been successful to date. One day, a user uploads an altered version of the AVM that is able to bypass the technical measures. The material is capable of being accessed in Australia, but no other user views the material and it is never reported. Until the time that the service provider becomes aware of the material, it may be taken to have been reckless with regard to that material being accessible.*

While it is entirely possible that the AVM Act intended to cover situations similar to the above scenarios, industry would benefit from guidance as to whether and how obligations under the AVM Act may realistically change over time.

The issue identified above also has an impact on s 474.35, regarding notices issued by the eSafety Commissioner. In the second-reading speech for the AVM Act, it was noted that "*once a notice has been given, that service will have no excuse for failing to comply with its obligation to remove that material as expeditiously as possible*". While such a proposition may be logical with respect to an initial notice and removal of content, it is unclear whether the presumption of recklessness continues to apply on an ongoing basis to all further copies and derivative versions of that material, especially and including those that escape industry-leading technological detection. If that were to be the case, Microsoft would suggest that this would be an adverse and unworkable outcome for industry participants.

Left unchecked, the issue described above will only increase over time as the amount of data (lawful and unlawful) shared online increases exponentially each year. As industry participants seek to comply with not only the AVM Act but also similar obligations under the OSA, realistically ringfencing compliance obligations will be critical to ensuring industry participants can effectively support online safety outcomes.

Recommendations regarding takedown scope over time

Microsoft recommends the Commonwealth Government provide clear guidance as to the treatment of AVM over time.

Additional areas of ambiguity

Microsoft has identified several other areas of ambiguity which we believe to be material to the Inquiry and its Terms of Reference:

> **Derivative material and copies**: The AVM Act should clearly address the possibility and impact of derivative versions and copies of AVM. While it may be assumed that the designation of certain

material as AVM would flow to its derivative forms, such a treatment has not been the case under similar legislative provisions.

Content classification decisions made by the Classification Review Board, and related takedown powers soon to commence under the OSA (previously under the *Broadcasting Services Act 1992* (Cth)), have been known to apply exclusively to specific copies of material and not others (despite having an identical quality in substance). For example, when the Classification Review Board classified the film 'Barnens O' as Refused Classification, the decision applied only to one particular copy of the film known as 'ACMA INV-0000-3781' hosted on a particular website, and not the film itself.

As this potential exists in sufficiently related legislative schemes, Microsoft sees it as critical for similar questions to be clarified in the context of the AVM Act.

"Expeditiously" and "reasonable time": Overall, the AVM Act is framed in fairly urgent terms, and there is a level of logic to this given the significant online harm it intends to combat. However, the time-based language used is still ambiguous and has been the cause of significant industry confusion.

Under s 474.33, service providers must notify the AFP "within a reasonable time after becoming aware of the material", and under s 474.34, service providers commit an offence if they do not ensure the "expeditious" removal of the material.

As explored above at 30, these unclear and somewhat subjective, yet nonetheless rapid timeframes, create an urgency that does not afford appropriate time to make the determinations currently expected of industry participants. If the AVM Act continues unamended and without further legislative clarification, we recommend that the Commonwealth Government provide further guidance for industry surrounding how these timeframes are to work in practice. Such guidance should take into account the real-world settings and practical challenges brought to the fore during this Inquiry.

Criminal sanctions: Microsoft questions the appropriateness of, the criminal sanctions contained in the AVM Act We do not see how criminal sanctions (including financial penalties and imprisonment) are proportionate penalties in this space, particularly in circumstances where a service provider is doing all it can feasibly do and applying significant resources in complying with the AVM Act. As noted earlier, the existence of these disproportionate criminal penalties risks having a chilling effect on tech investment in Australia. We encourage the PJCLE to consider how these penalties might be more appropriately aligned with the penalties related to other illegal content offences, and the potential effect of the overall regulatory regime on Australia's growing digital economy.

Hosting services: Microsoft continues to be deeply concerned that the AVM Act has disproportionate implications for hosting services and their customers, due to the technical and practical limitations that exist for these providers in responding to AVM Act obligations. Hosting service providers do not have the same ability to monitor and moderate discrete items of content as the other services in scope for AVM. Indeed, the Australian Government has subsequently recognised this in the OSA, with different requirements for hosting service providers.

In most cases, enterprise service providers such as Microsoft do not have the technical ability to identify, locate or remove discrete instances of content that are stored on their customer's services. Because the enterprise customer is closest to the ultimate end users of a service, that customer will be more likely to become aware of AVM and will be best positioned to remove or disable access to that content. Secondly, should AVM appear on a service, oftentimes the hosting service provider's only option will be to suspend service to that enterprise customer as a whole. This is a blunt tool, and risks having a disproportionate response on the enterprise customer Taken

alongside the other issues that we have raised, we strongly urge the Committee to recommend removing hosting services from the scope of the AVM Act.

3. Conclusion

Our intention in this submission is to bring to the Inquiry's attention the fact that the regulatory settings in place prior to the commencement of the AVM Act have changed so significantly so as to obviate the need for the AVM Act. We believe the objectives of the AVM Act can be – and are already being - met through other measures. If the Act is retained, we offer several suggestions to amend the Act to address some of the challenges that have become evident since its enactment. The Committee has the opportunity to take a considered and consultative approach to the AVM Act.

Microsoft again thanks the PJCLE for the opportunity to provide this submission to the Inquiry. We are available to discuss this submission with the PJCLE and would welcome the opportunity to appear before the Committee to provide further testimony on these important issues. We would also welcome the opportunity to discuss any proposed amendments with the Committee, including how it might give effect to our recommendations.