Bond University

Bond Law Review

Volume 33 Issue 1

2021

The Sharing of Abhorrent Violent Material Act: The Realities and Implications of Australia's New Laws Regulating Social Media Companies

Jasmine Valcic
Bond University

The *Sharing of Abhorrent Violent Material Act*: The Realities and Implications of Australia's New Laws Regulating Social Media Companies

JASMINE VALCIC*

On 15 March 2019, a Facebook Live video was broadcast from Christchurch, New Zealand, documenting a terror attack which resulted in the death of fifty-one people. This attack highlighted a weakness in social media protections and a gap in legislation globally. In response, the Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 *(Cth) seeks to make internet service and social media providers accountable for the removal of abhorrent and violent content. This legislation sent waves through the international community, attracting criticisms for its fast adoption, perceived unrealistic obligations and harsh penalties, as well as its broad extraterritorial reach. This article will explore these criticisms. It asks, how is the Act exercising extraterritorial jurisdiction? Is there an unrealistic burden created? And if there is a breach, who will be charged? The article concludes that if these challenges are not adequately addressed, the Act will not only fail to achieve its goal of reducing the accessibility of abhorrent violent material, but will also pose a serious threat to the protection of human rights.*

I Introduction

On Friday 15 March 2019, Brenton Tarrant started a stream on Facebook Live.[1] The video, lasting 16 minutes and 55 seconds, documented Tarrant opening fire on people praying at the Al Noor Mosque. Fifty-one people died that day. As expressed by New Zealand's Prime Minister, Jacinda Ardern, the terrorist attack "was designed to be broadcast on the internet".[2] There were up to 200 viewings during the live stream, totalling 4000 views before the original

* LLB Candidate, Bond University.

1 A video feature for Facebook designed to allow broadcasters to interact with viewers in real time.

2 Jacinda Arden, 'How to Stop the Next Christchurch Massacre', *The New York Times* (Opinion Post, 11 May 2019) <https://www.nytimes.com/2019/05/11/opinion/sunday/jacinda-ardern-social-media.html>.

video was removed.[3] Despite these views, the video was only reported by a user to Facebook 12 minutes after the live broadcast ended. Facebook only removed the video 41 minutes after that first user report, by which time copies had spread across the internet to sites such as 8kun,[4] and YouTube.[5] This attack highlighted a weakness in social media protections and a gap in legislation globally. In response, over 48 states and eight online service providers supported the Christchurch Call,[6] which asks for governments internationally to take steps "to eliminate terrorist and violent extremist content online".[7]

As a direct reaction to the Christchurch Massacre, on 3 April *the Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill 2019* was introduced to the Senate and by 5 April the Bill had received royal assent. The *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (Cth) ('*SAVM Act*') seeks to make internet service and social media providers accountable for the removal of content. The Act was met with criticism from industry professionals, international organisations and other state governments.[8] Key concerns were the lack of industry consultation during the drafting process, the ambiguities in the law, and the 'unrealistic' expectations the law imposes on internet service and social media providers.[9]

The article will be structured as follows. First, the nuances of the offences created under the *SAVM Act* will be explained, including the international response to the legislation. Then, the challenge of extraterritorial jurisdiction will be addressed in two parts looking at the exercise of prescriptive jurisdiction and of enforcement jurisdiction, concluding that extraterritoriality is not invoked in the prescriptive jurisdiction, but is likely to be required when enforcing the *SAVM Act*. This article will not address the issues involved in prosecutions under

---

3 Guy Rosen, 'A Further Update on New Zealand Terrorist Attack', *Facebook Newsroom* (Web Page, 20 March 2019) <https://about.fb.com/news/2019/03/technical-update-on-new-zealand/>.

4 Previously called 8chan, a user created and monitored message board, with minimal intervention from site administration.

5 Founders: New Zealand and France. Supporters: Australia, Canada, European Commission, Germany Indonesia, India, Ireland, Italy, Japan, Jordan, The Netherlands, Norway, Senegal, Spain, Sweden, United Kingdom, Argentina, Austria, Belgium, Bulgaria, Chile, Colombia, Costa Rica, Cyprus, Denmark, Finland, Georgia, Ghana, Greece, Hungary, Iceland, Ivory Coast, Kenya, Latvia, Lithuania, Luxembourg, Maldives, Malta, Mexico, Mongolia, Poland, Portugal, Romania, South Korea, Slovenia, Sri Lanka, Switzerland, UNESCO, Council of Europe.

6 Amazon, Daily Motion, Facebook, Google, Microsoft, Qwant, Twitter and YouTube.

7 New Zealand Ministry of Foreign Affairs and Trade, 'The Call', *Christchurch Call* (Web Page) <https://www.christchurchcall.com/call.html>.

8 See the 'Letter from Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression; and Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms while Countering Terrorism to Minister for Foreign Affairs Ms Marise Payne, 4 April 2019' and Sam Shead, 'YouTube, Facebook, Twitter targeted by strict new social media laws in Australia. Here's what it means', *Business Insider Australia* (4 April 2019).

9 Ibid.

extraterritorial jurisdictions as the challenges are not specific to the *SAVM Act* but affect all acts with an extraterritorial reach.

This article will then explore the challenges in determining who should be charged in the instance of a breach. Often if content is shared on one platform it will be present on numerous other social media platforms and, to further add to the complexity, most social media providers outsource their content moderation to third parties and have subsidiaries in many different countries.

Lastly, the article will discuss the burden the *SAVM Act* places on social media companies, and whether that burden is possible to uphold. Current Artificial Intelligence ('*AI*') technology was unable to keep up with individuals sharing the Christchurch Massacre footage and, despite the efforts of the social media providers, copies of the footage were still available weeks later.[10] Further, there is a risk that in their efforts to adhere to the *SAVM Act*, the freedom of expression and information may be negatively impacted. Overall, this article will conclude that the *SAVM Act*, while noble in its aspirations, is unenforceable and creates unrealistic expectations for social media and internet companies, with currently existing technological capabilities.

II The *Sharing of Abhorrent Violent Material Act*

The *SAVM Act* commenced on 6 April 2019, as a direct response to the events of the Christchurch Massacre. The purpose of the *SAVM Act* is to "ensure that online platforms cannot be exploited and weaponised by perpetrators of violence".[11] The legislation also sought to "send a clear message that the Australian government expects the providers of online content and hosting services to take responsibility for the use of their platforms to share abhorrent violent material".[12] The Act introduced two new offences, one for the failure to report abhorrent violent material ('*AVM*'), and the other for the failure to remove AVM.

A. *Section 474.33 Failure to Report*

Section 474.33 is an offence for failing to notify the Australian Federal Police ('*AFP*') within a *reasonable time* about material relating to *abhorrent violent conduct* occurring or that occurred in Australia. The offence applies to internet providers and hosting and content providers.

---

10 Donie O'Sullivan, 'Seven Weeks Later, Videos of New Zealand Attack Still Circulating on Facebook and Instagram', *CNN Business* (Web Page, 2 May 2019) <https://edition.cnn.com/2019/05/02/tech/new-zealand-video-instagram-facebook/index.html>.

11 Explanatory Memorandum, Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill 2019 (Cth) 2.

12 Commonwealth, *Parliamentary Debates*, House of Representatives, 4 April 2019, 1850 (Christian Porter, Attorney General).

The *SAVM Act* has not defined *'reasonable time'*, however, the Fact Sheet on the *SAVM Act* published by the Attorney-General's Department states that definition will depend on the unique circumstances in each case and that a wide range of factors will contribute to the determination of 'reasonableness'.[13] *Abhorrent violent conduct* is defined in the *SAVM Act* as conduct where the person engages in a terrorist act, murders or attempts to murder, tortures, rapes or kidnaps another person.[14] Although the abhorrent violent conduct must have occurred in Australia, it is immaterial whether the content or hosting service is provided within or outside Australia.[15] AVM is audio, visual or audio-visual material of abhorrent violent conduct produced by a perpetrator or their accomplice.[16]

To be prosecuted under s 474.33 the provider must:

1. Have been aware that their service can be used to access particular material; and
2. Have reasonable grounds to believe that the material was abhorrent violent material; and
3. Have had reasonable grounds to believe that the relevant material was occurring or had occurred in Australia.[17]

The offence does not extend to ignorance or negligence. For example, if a content provider is unaware of the AVM, they cannot be charged for a failure to report under s 474.33. Further, the offence does not apply if the provider reasonably believes that details of the material are already known to the Australian Federal Police.[18] Note the evidential burden lies on the defendant to prove a reasonable belief that the Australian Federal Police were already aware of the conduct.[19] The offence carries 800 penalty units.

B. *Section 474.34 Failure to Remove*

Section 474.34 creates an offence for content and hosting services which fail to remove access to AVM *expeditiously* where that material is reasonably capable of being accessed in Australia.[20] This offence is not applicable to internet service providers. The offence has

---

13 Attorney-General's Department, *Sharing of Abhorrent Violent Material Act Fact Sheet*, July 2019 <https://www.ag.gov.au/Crime/federal-offenders/Documents/AVM-Fact-Sheet.pdf> ('*Fact Sheet*').

14 *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (Cth) s 474.32 ('*SAVM Act*').

15 Ibid s 474.33(2).

16 Fact Sheet (n 13).

17 Ibid.

18 *SAVM Act* (n 14) s 474.33(3).

19 Ibid.

20 Ibid s 474.34.

extraterritorial reach, applying to all providers irrespective of whether the content service is provided within or outside Australia.[21]

However, the offence only applies where the material is reasonably capable of being accessed within Australia. Content service is defined as a social media service or a designated internet service, within the meaning of the *Enhancing Online Safety Act 2015* (Cth).[22] Hosting service has also been given the same meaning as in the *Enhancing Online Safety Act*.[23]

Recklessness is the fault element for the offence.[24] There are two elements in determining if there is the breach:

1. Whether the material is accessible through the service; and
2. Whether the material is AVM.[25]

The provider is only liable where they were aware of a substantial risk that their platform can be used to access AVM and, having regard for the circumstances known to the provider, it was unjustifiable to take the risk and intentionally failed to remove the AVM expeditiously.[26] The offence carries two penalties, one for the individual and one for the body corporate. An offence under this section by an individual is punishable by imprisonment for a period of not more than three years, or a fine of up to 10,000 penalty units, or both.[27] If the offence is committed by a body corporate, the penalty is a fine no greater than 50,000 penalty units, or 10% of the annual turnover of the body corporate during the period of 12 months ending at the end of the month in which the conduct constituting the offence occurred.[28]

---

21 Ibid s 474.34(6).

22 Ibid s 474.30; *Enhancing Online Safety Act 2015* (Cth), s 9 – A social media service is an electronic service that's sole or primary purpose is to enable online social interaction between two or more end-users, the service allows end-users to link to, or interact with, some or all of the other end-users, the service allows end-users to post material service, and fulfills any other conditions as set out in the legislative rules. Section 9A – A designated internet service means a service that allows end-users to access material using an internet carriage service, or a service that delivers material to persons having equipment appropriate for receiving that material, where the delivery of the service is by means of an internet carriage service (not including a social media service, relevant electronic service, an on-demand program service, or a service specified by the minister by legislative instrument).

23 *SAVM Act* (n 14) s 474.30; *Enhancing Online Safety Act 2015* (Cth) s 9C – If a person (the *first person*) hosts stored material that has been posted on a social media service, relevant electronic service, or designated internet service AND the first person or another person provides a social media service, relevant electronic service or a designated internet service on which the hosted material is provided the hosting of the stored material by the first person is taken to be the provision by the first person of a *hosting service*.

24 SAVM Act (n 14) s 474.34(4).

25 Fact Sheet (n 13).

26 Ibid.

27 *SAVM Act* (n 14) s 474.34(9).

28 Ibid s 474.34(10).

C. *Limits on the Offences*

Section 474.34 does not apply if the accessibility of the material is necessary for enforcing or monitoring the compliance with a law of Australia or a foreign country.[29] There are also defences for journalists, scientific, medical, academic or historical research, and 'artistic work'.[30] The evidentiary burden for all defences lies on the defendant.

Additionally, both offences only apply to footage of abhorrent violent conduct filmed by the perpetrator or their associates and do not encompass footage captured by innocent bystanders.[31] The offences do not apply to the extent that they would infringe any constitutional doctrine of implied freedom of political communication.[32]

Any proceedings under s 474.33 cannot be commenced without the Attorney-General's written consent if the conduct constituting the alleged offence occurs wholly in a foreign country, and the person alleged to have committed the offence is neither an Australian citizen nor a body corporate incorporated in Australia.[33]

All proceedings for an offence against s 474.34 require the Attorney-General's written consent.[34] However, an individual may be arrested, charged or remanded in custody in connection with either offence prior to receiving the consent.

Under sections 474.35 and 474.36, the eSafety Commissioner may issue a written notice stating that AVM was being hosted or could be accessed on the content or hosting service. This notice does not mean an offence has been committed; it simply puts the provider on notice that their service can be used to access AVM. The notice creates a rebuttable presumption, in relation to future prosecution, that the provider was reckless.[35]

III National and International Response

The international response to the *SAVM Act* was critical. The United Nations Special Rapporteurs for Freedom of Expression and Countering Terrorism expressed their concerns to the Government that the approach to the *SAVM Act*, particularly the haste in presentation and adoption, as well as that key elements of the law unduly interfere with Australia's obligations under international human rights law.[36]

---

29 Ibid s 474.47(1).
30 Ibid s 474.47.
31 Fact Sheet (n 13).
32 *SAVM Act* (n 14) s 474.38.
33 Ibid s 474.42(1).
34 Ibid s 474.42(3)
35 Fact Sheet (n 13).
36 Letter from Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression; and Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms while Countering Terrorism to Minister for Foreign Affairs Ms Marise Payne, 4 April 2019, 1 ('*Letter from Special Rapporteur*').

There was further industry outcry, Managing Director of the Digital Industry Group Inc (DIGI),[37] Sunita Bose, criticised the *SAVM Act* for its lack of meaningful consultation, given the complexity of the problem.[38] Additionally, Prime Minister Ardern said New Zealand would not follow Australia's hard-line response.[39]

It is interesting to note that Germany instituted a similar law in 2017, commonly known as *NetzDG*,[40] which requires large social media companies to proactively enforce German speech laws on their platforms.[41] *NetzDG* was met with similar international criticism.[42] The German government issued its first fine under *NetzDG* in July 2019 to Facebook. The company was required to pay €2 million for under-reporting illegal activity on its platforms in Germany; however, the company did complain that the law lacked clarity.[43] Further, on 18 April 2019, the EU Parliament voted to fine internet firms up to four per cent of their turn over if they persistently fail to remove extremist content within one hour of being asked to do so by authorities.[44] However, at the time of writing, the content of the law has not been finalised.

Notably, there has been a shift in Facebook's approach to government regulation of social media, although the *SAVM Act* was not specifically mentioned, Mark Zuckerberg, founder and chief executive of Facebook, said: "I believe we need a more active role for governments and regulators […] I believe we need new regulation in four areas: harmful content, election integrity, privacy and data

---

37 DIGI represents Facebook, Google and Twitter.

38 Sam Shead, 'YouTube, Facebook, Twitter targeted by strict new social media laws in Australia. Here's what it means.', *Business Insider Australia* (Web Page, 4 April 2019) <https://www.businessinsider.com.au/youtube-facebook-twitter-targeted-by-strict-new-social-media-laws-in-australia-heres-what-it-means-2019-4>.

39 Jenna Lynch, 'Jacinda Ardern will not Follow Australia's Hard-line Response to Extremist Content', *Newshub* (Web Page, 19 July 2019) <https://www.newshub.co.nz/home/politics/2019/07/jacinda-ardern-will-not-follow-australia-s-hard-line-response-to-extremist-content.html>.

40 *Netzwerkdurchsetzungsgesetz* (NetzG) gives social media sites up to 24 hours after notification to remove "obviously illegal" content. A failure would result in fines of up to 50m Euros. NetzG was enacted to addressed hate speech and discrimination on social media.

41 'Germany Starts Enforcing Hate Speech Law', *BBC News* (Web Page, 1 January 2018) <https://www.bbc.com/news/technology-47135058>.

42 Amélie Heldt, 'Reading Between the Lines and the Numbers: An Analysis of the First NetzDG Reports' (2019) 8(2) *Internet Policy Review*.

43 'Germany Starts Enforcing Hate Speech Law' (n 41).

44 Foo Yun Chee, 'EU Parliament Votes to Fine Internet Firms for Not Removing Extremist Content Quicklty', *Reuters* (Web Page, 18 April 2019) <https://www.reuters.com/article/us-eu-parliament-extremist-content/eu-parliament-votes-to-fine-internet-firms-for-not-removing-extremist-content-quickly-idUSKCN1RT2CF>.

portability".[45] Zuckerberg went further to say "[i]nternet companies should be accountable for enforcing standards on harmful content".[46]

IV Extraterritorial Jurisdiction

There are three types of jurisdiction when exploring extraterritoriality. Prescriptive jurisdiction is the ability for the state to prescribe or legislate in respect of persons and conduct.[47] Enforcement jurisdiction is the capacity of the state to enforce the laws.[48] Adjudicative jurisdiction refers to the ability for the courts of that state to adjudicate and resolve disputes.[49] Here, although at first glance the two offences created in the *SAVM Act* seem to have extraterritorial jurisdiction, they may not have extraterritorial reach in prescriptive jurisdiction.

A. *Prescriptive Jurisdiction*

Both offences require a connection between the conduct and Australia. In the failure to report, there is a requirement for the conduct the AVM depicts to be occurring or have occurred within Australia.[50] To be liable under the offence of failure to remove, the AVM must be reasonably capable of being accessed within Australia.[51] These two clauses create a territorial nexus between the conduct constituting the offence and Australia. This would seem to suggest that objective territoriality applies, rather than an extraterritorial reach. The territoriality principle of jurisdiction is the most common and least controversial basis of jurisdiction, stemming from the founding principle of state sovereignty.[52]

Territoriality affirms that states have jurisdiction over the conduct which occurs within their territorial borders.[53] Objective territoriality refers to the state's jurisdiction over conduct which only partially occurs within that state's territory.[54] Danielle Ireland-Piper provides a clear illustration of objective territorial jurisdiction; a gun is fired in

---

45 Mark Zuckerberg, 'The Internet Needs New Rules. Let's Start in These Four Areas.', *The Washington Post* (Opinion Post, 31 March 2019) 3 <https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html>.

46 Ibid 6.

47 Alex Mills, 'Rethinking Jurisdiction in International Law' (2014) 84(1) *British Yearbook of International Law* 187, 195.

48 Gillian D Triggs, *International Law: Contemporary Principles and Practices* (LexisNexis Butterworths, 2nd ed, 2010).

49 Danielle Ireland-Piper, *Accountability in Extraterritoriality: A Comparative and International Law Perspective* (Edward Elgar, 2017) 21.

50 *SAVM Act* (n 14) s 474.33(1)(b).

51 Ibid s 474.34(3).

52 Ireland-Piper (n 49) 22.

53 Gerhard Kegel and Ignaz Seidl-Hohenveldern, 'On the Territoriality Principle in Public International Law' (1982) 5(2) *Hastings International and Comparative Law Review* 245 249.

54 Ireland-Piper (n 49) 22.

State A, the bullet crosses the border into State B, where it causes injury. Even though the conduct, the pulling of the trigger, took place in State A, the injury from the bullet occurred in State B.

As a result, State B may assert jurisdiction on the basis of objective territoriality.[55] With the *SAVM Act*, the depicted conduct in Australia and the accessibility of the AVM in Australia, are the 'injury' of the offences respectively, even though the conduct, failing to remove or report, occurred in another state.

The application of objective territorial jurisdiction in criminal law is affirmed in *Halsbury's Laws of Australia*: "A state may also assert jurisdiction where the effects of the alleged criminal act are felt within the territory, although the commission of the offence occurs elsewhere".[56] The concept was explored and applied by the High Court in the case of *Lipohar v R* ('*Lipohar*').[57] In *Lipohar* the appellants were tried in the Supreme Court of South Australia on a charge of conspiracy to defraud.[58] The development of the conspiracy was formed, and relevant steps were taken wholly outside of South Australia.[59] Further, none of the conspirators were residents of South Australia. However, the intended victim of the conspiracy was Collins Street Properties Pty Ltd, which was incorporated in South Australia. The appellants appealed to the High Court on the basis that the Supreme Court of South Australia did not have jurisdiction to try the offence. Dismissing the appeal, Gleeson CJ asserted that the fact the intended victim was a South Australian company, and the impact of the conspiracy would have been felt in South Australia created a sufficient nexus to justify territoriality.[60] Therefore, although ss 474.33(2) and 474.34(6) suggest an extraterritorial reach, in its prescriptive jurisdiction, the *SAVM Act* is merely invoking objective territorial jurisdiction.

However, although the prescriptive jurisdiction is not extraterritorial, the enforcement jurisdiction may be and as a result, could raise several issues regarding double jeopardy and conflicting international laws.

B. *Enforcement Jurisdiction*

Although a nexus to Australia is still present in the *SAVM Act*, the offences act extraterritorially in the enforcement jurisdiction as it regulates the conduct of non-Australian citizens outside of the Australian territorial border. In this way, the *SAVM Act* differs from

---

55 Ibid 23.

56 LexisNexis, *Halsbury's Laws of Australia,* 215 Foreign Relations, '3 Territory and Jurisdiction' [215-380].

57 (1999) 168 ALR 8 ('*Lipohar*'). First explored regarding interstate extraterritorial jurisdiction in *Ward v R* (1980) 29 ALR 175.

58 Ibid 38-40.

59 The scheme involved activity in Indonesia, Thailand, Queensland and Victoria.

60 *Lipohar* (n 57) 38-40.

Australia's previous exercise of extraterritorial jurisdiction in criminal law. For example, one of Australia's most cited criminal offences with international reach is the *Crimes Legislation Amendment (Sexual Offences Against Children) Act 2010* (Cth) ('*SOAC*'). The *SOAC* regulates conduct which occurs beyond Australia's territorial borders; s 272.8 criminalises sexual intercourse with a child under the age of 16 outside of Australia.[61] However, unlike the *SAVM Act*, the *SOAC* exercises the Active Nationality Principle.[62] Pursuant to s 272.6, only Australian citizens, residents of Australia, body corporates incorporated in Australia, or other body corporates that carries on its activities primarily in Australia, can be prosecuted under the Act.[63] The extraterritorial nature of the *SOAC* differs drastically from the *SAVM Act*, which, according to ss 474.33(2) and 474.34(2), applies to content and hosting services provided outside Australia.[64]

Division 15 of the *Criminal Code Act 1995* (Cth), (the '*Criminal Code*'), outlines the 'categories' of extended geographical jurisdiction for offences under the *Criminal Code*. Interestingly, the *SAVM Act* is not expressly assigned to one of these categories. An example of this division being applied is the *Cybercrime Act 2001* which, under s 476.3, falls within Category A. Category A is defined in s 15.1 of the *Criminal Code*, which states that offences within the category apply to conduct constituting the alleged offence which occurs wholly outside of Australia and a result of the conduct occurs wholly or partly in Australia.[65]

On face value, this would make the extraterritorial reach similar to the *SAVM Act*. However, a clear distinction arises in the available defences. Pursuant to s 15.1(2), a person does not commit an offence in Category A if the alleged offence is a primary offence, the conduct constituting the offence occurs wholly in a foreign country, the person is neither an Australian Citizen nor a body corporate incorporated under Australian law, and there is not a law in the foreign country where the conduct occurs which corresponds to the Australian offence.[66] In short, conduct in a foreign country by a foreign national would only constitute an Australian Category A offence if the foreign country has a corresponding law covering the Australian offence.

If the *SAVM Act* were classified as Category A, this defence would limit the application of the act significantly, as the only state with a potentially sufficient corresponding law is Germany, with its *NetzDG*

---

61 *Crimes Legislation Amendment (Sexual Offences Against Children) Act 2010* (Cth) s 272.8 ('*SOAC*').

62 Active nationality principle refers to a state's jurisdiction over the conduct of its citizens overseas. It is the strongest basis for direct extraterritorial jurisdiction.

63 *SOAC* (n 61) s 272.6.

64 *SAVM Act* (n 14).

65 *Criminal Code Act 1995* (Cth) s 15.1(b) ('*Criminal Code*').

66 Ibid s 15.2.

legislation. However, as the *SAVM Act* has not been assigned a category, these defences do not apply. It is clear the *SAVM Act* is intended to have a much broader extraterritorial enforcement jurisdiction. This more extensive jurisdiction may raise issues surrounding conflicting international laws and double jeopardy, which would complicate or even prevent the effective enforcement of the Act.

1 *Conflicting International Laws*

There are two scenarios in which an issue of conflicting international laws could arise in the exercise of this broad extraterritorial enforcement jurisdiction. First, a situation in which the requirement of one state's law would require the contravention of a second state's law. Second, a situation in which the enforcement of one state's law would contravene a second state's law. The first scenario is significantly more straight forward in its application to the *SAVM Act*. For example, the DIGI director expressed concern that the Act would require internet, content and hosting providers registered in the United States who operate in Australia to breach United States law.[67]

Specifically, there are laws in the United States, where all DIGI founding members are headquartered, that forbid companies from sharing certain types of information, particularly content data, with law enforcement agencies outside of the United States.[68] There is a clear conflict between this United States law and the requirements under s 474.33 to inform the Australian Federal Police of AVM depicting conduct that is reasonably believed to be or have taken place in Australia. This is only one example of potential tension. As the *SAVM Act* applies to all hosting and content service providers, irrelevant of where they are based or incorporated, many providers will find themselves subject to the *SAVM Act* as well as the laws of the state in which they are incorporated or registered. This conflict of national laws is also not easily resolved and may hinder the extraterritorial enforceability of the *SAVM Act*.[69]

The second instance of a conflict between laws of different states is when the enforcement of one law may infringe or contravene a second state's law. This scenario was explored by a French case in 2000; *L'Union Des Etudiant Juifs De France Et La Ligue Contre Le Racisme Et L'Antisemitisme v Yahoo! France* ('*LICRA Case*'),[70] and the

---

67 Ariel Bogle, 'Laws Targeting Terror Videos on Facebook and YouTube 'Rushed' and 'Knee-Jerk', Lawyers and Tech Industry Say', *Australian Broadcasting Corporation News* (Web Page, 4 April 2019) <https://www.abc.net.au/news/science/2019-04-04/facebook-youtube-social-media-laws-rushed-and-flawed-critics-say/10965812>.

68 Ibid.

69 The potential solution for these conflicts is explored in the area of research known as 'Conflict of Laws', and there are various methodologies employed to determine which law should prevail depending on relevant parties and circumstances.

70 France, T.G.I. Paris, May 22, 2000 ('*LICRA Case*').

subsequent United States ('*US*') case brought by Yahoo!, *Yahoo!, Inc. v La Ligue Contre le Racisme et L'Antisemitisme* ('*Yahoo! Case*').[71] In 2000, two French student organisations brought a civil action against internet service provider Yahoo!, for breaching the French Penal Code. Article R645-1 of the French Penal Code criminalises the display, exchange or sale of Nazi or Third Reich paraphernalia and memorabilia.[72] Yahoo! is an internet service provider incorporated under the laws of Delaware, and operated principally in California.[73] Yahoo!'s auction site allowed the posting of items that are illegal in France, including Nazi paraphernalia.[74] LICRA and UEJF,[75] the two student organisations, sent a cease and desist letter to Yahoo!'s headquarters in California, detailing that the sale of the Nazi paraphernalia through the auction site violated French Law.

LICRA and UEJF gave Yahoo! eight days to prevent these sales, or they would take legal action. After the eight days expired, LICRA and UEJF filed a complaint against Yahoo! in the *Tribunal de Grande Instance de Paris*.[76] This case was novel in its exploration of the global nature of the internet and extraterritorial reach concerning the internet. The French Tribunal found that there was a sufficient nexus between the conduct of Yahoo! in allowing the auctioning of Nazi paraphernalia and the French jurisdiction in that French citizens have access to approximately 1,000 offending materials on Yahoo.com.[77] In May 2000, the French Tribunal ordered Yahoo! to:

1. Eliminate French Citizens' access to any Nazi or Third Reich paraphernalia or memorabilia on the Yahoo.com auction site; and
2. Eliminate French citizen's access to pages on Yahoo.com that display text, extracts, or quotations from *Mein Kampf,* Hitler's autobiography and other relevant texts; and
3. Post a warning on Yahoo! France stating that searches on Yahoo.com could lead to sites containing material prohibited by R645-1 of the French Criminal Code and that the viewing of such material could result in legal actions; and

---

71 169 F. Supp. 2d. ('*Yahoo! Case*').
72 *Penal Code* (France) Article R654-1 (2001).
73 Elissa A Okoniewski, 'Yahoo!, Inc. v. LICRA: The French Challenge to Free Expression on the Internet' (2002) 18(1) *American University International Law Review* 295, 311.
74 Ibid.
75 La Ligue Contre Le Racisme Et L'Antisemitisme and L'Union Des Etudiant Juifs De France.
76 Okoniewski (n 73) 313.
77 These materials included copies of Adolf Hitler's *Mein Kampf*; *The Protocols of the Elders of Zion* and purported 'evidence' of the non-existence of the Holocaust gas chambers.

4. To remove from browser directories accessible in France index headings entitled "negationists".[78]

Following this order, Yahoo! filed a complaint in the US District Court for the Northern District of California. Yahoo! sought a declaratory judgement that under the laws of the United States, the French Court's orders were neither enforceable nor recognisable.[79] Yahoo! asserted that they could only comply with the French order by completely banning all Nazi-related goods from the site,[80] and that this ban would constitute an infringement on its First Amendment rights.[81]

The underlying question was whether another state could regulate speech within the United States without violating the First Amendment on the grounds the speech is accessible via the internet in that state. The *SAVM Act* is likely to face this same question in its enforcement. Overall, the Court granted the declaration that the First Amendment precludes enforcement of the French order within the United States. The Court, in its reasoning, stated that speech spoken in the United States could not be regulated by a foreign nation just because it could be heard there.[82] The *SAVM Act* would likely be met with a similar response from foreign courts, if not the same response. This case study is particularly pertinent as most large social media and content providers are based in the United States. The impact on the extraterritorial enforceability of the offences under the *SAVM Act* would be severely hindered if the United States declared each judgement unenforceable domestically.

2 *Double Jeopardy*

The other issue raised by the extraterritorial nature of the *SAVM Act* is international double jeopardy. The principle of *non bis in idem*, known as double jeopardy, protects individuals from repeated prosecution for offences arising out of a singular event.[83] Although the principle is well-founded in the domestic law of many states, there is significant ambiguity in its applications across multiple sovereignties. Under the current frameworks governing international laws, a national prosecution enforcing a national law does not bar successive

---

78 Negationists is a French term used to refer to theories and writings that question the existence of the Holocaust or elements of it.

79 Okoniewski (n 73) 316.

80 *Yahoo! Case* (n 71) 1185-1186.

81 The 1st Amendment of the US Constitution protects the freedom of speech, religion and expression. ("*Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.").*

82 *Yahoo! Case* (n 71) 1194.

83 Michele N Morosin, 'Double Jeopardy and International Law: Obstacles from Formulating a General Principles', 64 *Nordic Journal of International Law* 261, 261.

prosecutions by other states with national jurisdiction over the crime in question.[84]

Article 14(7) of the *International Covenant on Civil Political Rights* is a general provision of double jeopardy,[85] but was interpreted by the Human Rights Committee as only limiting second trials within a single jurisdiction.[86]

As a result, if the conduct is interpreted to contravene both *NetzDG* and the *SAVM Act*, or other similar laws which may arise in additional states, the content or hosting service may face multiple prosecutions in different states for the same action. This raises significant concerns regarding justice and fairness. The concerns of double jeopardy are particularly pertinent with the *SAVM Act* as the nexus is that the content is accessible within Australia; as the internet is truly international, the extraterritorial reach is potentially global, as it would be for *NetzDG* or other similar laws.

Overall, the *SAVM Act* has extraterritorial reach in its enforcement jurisdiction. As aforementioned, the Australian Government will face significant international barriers to the effective enforcement of the *SAVM Act*; however, if it is enforced, there may be considerable concerns as to justice and fairness internationally.

V Whom Do We Prosecute?

If Australia can justify the extraterritorial reach of the *SAVM Act* to the international community, the next challenge the Australian Government will face is determining who to charge for breaches under the offences. Three key issues complicate this determination. First, the nature of the internet is not one of isolation; if the content is uploaded on one platform, it is likely to be accessible on numerous other social media platforms in mere minutes. Second, it would arguably be easier to charge the locally registered subsidiaries of the large companies; however, often, these subsidiaries have minimal assets and are not responsible for content moderation. Last, the majority of social media providers outsource their content moderation to third parties and, in this instance, it may be challenging to determine who is liable and the extent to which they are accountable. Each of these three concerns will be addressed separately. However, this article will not solve the issues, nor does it purport to declare definitively who the correct actors to charge under the offences are.

---

84 Anthony J Colangelo, 'Double Jeopardy and Multiple Sovereigns: A Jurisdictional Theory', 86(4) *Washington University Law Review* 769, 797.

85 *International Covenant on Civil and Political Rights*, opened for signature 16 December 1996, 999 UNTS 171 (entered into force 23 March 1976).

86 Morosin (n 83) 262. See also *AP v Italy No 204/1986*.

A. *Multiple Platforms*

Unfortunately, the content which the *SAVM Act* is attempting to remove from the internet is often 'viral content'. Most definitions of viral content include factors of spreading rapidly and very widely.[87] Experts are yet to come to a consensus on the exact numbers of views in a set amount of time required for content to be defined as viral.[88] Jonah Berger discusses 'virality' in terms of "how contagious something is, or how likely it is to be shared given exposure".[89] When content spreads widely and quickly, it is more often than not shared on several different platforms. The footage from the Christchurch Massacre was shared on YouTube, Facebook, Twitter, Instagram and Reddit, and 8kun. Much of the content captured by the *SAVM Act* is material which tends to go viral. Violent content often goes viral; however, experts have not yet reached a consensus as to why. Theories refer to violence in sport, living vicariously through others, the 'outrageousness' of the content, or perhaps that individuals just 'crave' violence.[90]

A clear example of the viral nature of the content the *SAVM Act* regulates is the tragic murder of Bianca Devin. In July 2019, 17-year-old Bianca Devin was allegedly murdered by Brandon Andrew Clark, who then shared photos of the murder on his Instagram Story, which included a graphic image of the victim sitting in an SUV with her neck severed.[91] This content undoubtedly falls within the definition of abhorrent violent conduct, specifically within the category of material depicting the murder of another person.[92] Although the content was first shared on Instagram, it was quickly dispersed to Discord and 4chan. The images could still be found across the internet days later.[93] It is likely that under the *SAVM Act*, Instagram, Discord and 4chan, as well as other fringe messaging boards could be charged with a failure to remove AVM.

However, the issue raised is if the content is shared across so many platforms, who should or would the Government be charging first, or at all? If the Attorney-General decides to charge all platforms in breach

---

87 Robert Wynne, 'There Are No Guarantees or Exact Statistics for Going Viral', *Forbes* (Web Page, 9 March 2018) <https://www.forbes.com/sites/robertwynne/2018/03/09/ there-are-no-guarantees-or-exact-statistics-for-going-viral/#13a14b085e8c>.

88 Ibid.

89 Ibid.

90 Kate Wheeling, 'Why Do Violent Videos Go Viral?', *Pacific Standard* (Web Page, 14 June 2017) < https://psmag.com/environment/violent-videos-in-atlanta-schools>.

91 E J Dickson, 'A 17-Year-Old Girl Was Murdered. How Did Photos of Her Death Go Viral?', *Rolling Stone* (Web Page, 15 July 2019) <https://www.rollingstone.com/culture/ culture-news/bianca-devins-murder-brandon-andrew-clark-858874/>.

92 *SAVM Act* (n 14) s 474.32(1)(b).

93 Queenie Wong, 'Instagram's Dark Side: Grisly Photos of Teen's Slaying Spread on Social Media', *net* (Web Page, 15 July 2019) <https://www.cnet.com/news/instagrams-dark-side-grisly-photos-of-teens-slaying-spread-on-social-media/>.

for the singular incident, it would likely be a significant drain on public funds.

Further, the court system would become inundated with prosecutions under the *SAVM Act*. There are many lines of reasoning the Attorney-General could take in deciding who to prosecute, for example, the platform which the content was first accessible, or the largest platform, or the platform with the highest number of views of the content. Overall, the Attorney-General would have to determine the best platform to prosecute for a breach, if the Act is going to be enforced. Interestingly, although the murder of Bianca Devin occurred after the enactment of the *SAVM Act*, there has been no announcement of the Attorney-General considering prosecution under the *Act*.

B. *Local Subsidiaries*

At first glance, one might think that it would be most efficient to prosecute the local subsidiaries of the more substantial parent content and hosting service providers. For example, Facebook Inc's local subsidiary is Facebook Australia Pty Ltd, and Google Inc's is Google Australia Pty Ltd.[94] Prosecuting the Australian registered subsidiaries would potentially bypass some of the extraterritorial enforcement issues. Further, it would likely be easier to obtain evidence, run court hearings and other required procedures.[95] However, prosecuting the local subsidiaries under the *SAVM Act* may not be a viable option for several reasons.

First, often the locally registered subsidiaries of the larger social media providers have significantly smaller assets than those of their parent company. For instance, the sales revenue for Facebook Australia Pty Ltd was $579,650 in 2018,[96] whereas the revenue for Facebook Inc was $55 billion.[97]

The penalty for a body corporate for a breach of s 474.34 is a fine of no greater than the following: 50,000 penalty units, and 10% of the annual turnover of the body corporate for 12 months ending at the end of the month in which the conduct constituting the offence occurred.[98] As of July 2017, 50,000 penalty units is equivalent to $10,500,000, which is more than 18 times the revenue of Facebook Australia. 10%

---

94 Note that this challenge is further complicated as Instagram is a subsidiary of Facebook Inc., and YouTube is a subsidiary of Google Inc., and Google's parent company is Alphabet Inc.

95 There are many logistical challenges with exercising extraterritorial jurisdiction such as local law enforcement co-operation, making all parties available in person for the court hearings etc.

96 'Facebook Australia Pty Ltd – Australian Company Profile', *IBIS World* (Web Page) <https://www.ibisworld.com.au/australian-company-research-reports/information-media-telecommunications/facebook-australia-pty-ltd-company.html>.

97 Facebook, 'Facebook Reports Fourth Quarter and Full Year 2018 Results', *Facebook Investor* (Web Page, 30 January 2019) <https://investor.fb.com/investor-news/press-release-details/2019/Facebook-Reports-Fourth-Quarter-and-Full-Year-2018-Results/default.aspx>.

98 *Crimes Act 1914* (Cth) s 4AA.

of the turnover of Facebook Australia would only be approximately $57,000. A fine this small would hardly incentivise a parent company with the revenue of Facebook Inc to invest in or implement stronger moderation technologies or practices. Therefore, to achieve the aim of holding companies accountable for the accessibility of AVM on their platforms, the Attorney-General should prosecute the parent company. However, as aforementioned, prosecuting the internationally registered parent company comes with its own host of issues.

A second challenge is the work and responsibilities of the local subsidiaries to the parent company and content service as a whole. In most instances, the locally registered subsidiary has little control over content moderation. Instead, most of their work revolves around advertising, marketing and sales. For example, Google Australia Pty Ltd is involved in advertising and information management technology services along with marketing and assistance services relating to its web search engine.[99] Similarly, Facebook Australia does not have the authorisation to access user records, and claims it does not "control or operate the website". These parameters raise the question: if Facebook Australia does not have any powers concerning the control or operation of the website, and does not directly engage in content moderation themselves, can they be held criminally liable under the *SAVM Act*? Further, if the local subsidiary does not have the capability or authorisation to alter moderation methods, the prosecution of them is unlikely to change the overall approach to moderation by the parent company.

The third challenge is related to the potential penalty for individuals for breaches of offences under the *SAVM Act*. Pursuant to s 474.34(9) a failure to remove AVM expeditiously by an individual is punishable on conviction by imprisonment for a person of up to 3 years, or a fine of up to 10,000 penalty units, or both. If the locally registered subsidiary is not involved in content moderations, can any of the local employees or directors be individually liable under the *SAVM Act*? The complexity of internal content moderation further complicates this issue. To illustrate, the Internal Moderating Team of Facebook consists of approximately 15,000 employees in 20 separate locations.[100] Given the large number of individuals involved in content moderation, it would be difficult to determine who should be individually liable for a failure to remove the offending AVM expeditiously.

---

[99] 'Google Australia Pty Ltd – Australian Company Profile', *IBIS World* (Web Page) <https://www.ibisworld.com.au/australian-company-research-reports/information-media-telecommunications/google-australia-pty-ltd-company.html>.

[100] Ellen Silver, 'Hard Questions: Who Reviews Objectionable Content on Facebook – And Is the Company Doing Enough to Support Them?', *Facebook Newsroom* (Web Page, 26 July 2018) <https://about.fb.com/news/2018/07/hard-questions-content-reviewers/>.

At first glance, the prosecution of the local subsidiary would seem most efficient as it does not require the same reliance on extraterritorial jurisdiction. Upon further scrutiny, the prosecution of the Australian registered subsidiary would fail to achieve the overarching goal of the *SAVM Act*, which is to reduce the accessibility of the AVM in Australia.

C. *Third-Party Content Moderation*

To further complicate the determination of which actor should be prosecuted, most of the larger content and hosting service providers outsource their content moderation to third party companies. For example, Facebook uses a host of different companies to support their content moderation, including Cognizant, PRO Unlimited, Accenture, Arvato, and Genpact.[101] Between the companies, there are content moderation sites in India, Dublin, Germany, and across the United States.[102] Each of these organisations has thousands of employees in their content moderation teams. There is a possibility that the larger content and hosting service provider may attempt to pass the criminal liability down the line to these third-party moderation services.

They may argue that it was the moderation service's contractual responsibility to remove the offending material and, therefore, the moderation service's failure to remove it expeditiously. Alternatively, it could lead to civil litigation between the content or hosting service provider and the content moderation service. Further, as there is a wide range of moderation companies used by the social media platforms, there will be apparent difficulties in pinpointing which company was responsible for the moderation of the relevant AVM. Additionally, as a result of the viral nature mentioned above of AVM, there is a high chance several of the moderation companies would have come across the offending material.

The role of content moderation companies affects not only the liability at a company level but also at the individual employee level. This uncertainty is similar to the issues with the local subsidiaries of the content and hosting service providers. Under s 474.34(9) a failure to remove AVM expeditiously by an individual is punishable on conviction. If an employee of a third-party content moderation company fails to remove AVM expeditiously, are they criminally liable under s 474.34(9)? Further, the potential application of the *Corporations Act 2001* (Cth) ('*Corporations Act*') should be considered regarding the individual criminal liability of directors. Notably, the *Corporations Act*

---

[101] Queenie Wong, 'Facebook Content Moderation is an Ugly Business. Here's Who Does It', *cnet* (Web Page, 19 June 2019) <https://www.cnet.com/news/facebook-content-moderation-is-an-ugly-business-heres-who-does-it/>.

[102] Ibid.

does not have broad extraterritorial jurisdiction.[103] The Government and courts will have to address these questions, and the role and responsibility of third-party content moderation services, and their employees, under the *SAVM Act*, and its impact on the criminal liability of the content and hosting service providers for a successful prosecution under the Act.

As highlighted, there are several questions regarding where criminal liability falls and who should be charged under the *SAVM Act*, which need to be addressed before a prosecution can commence. First, the Attorney-General would need to choose which content and hosting services to prosecute for the failure to remove the offending AVM given the viral and interconnected nature of the internet.

Second, the Attorney-General would then have to decide whether to prosecute the local subsidiaries, the subsidiary responsible for the content moderation of the relevant AVM or the parent company of the content or hosting service. Last, the Attorney-General and the Courts need to determine the responsibility and potential liability of third-party content moderation services which are used by the major social media companies. Without answering these three key questions, it would be near impossible for the Attorney-General to proceed with a prosecution under the *SAVM Act*.

VI Impossible Burden

The final hurdle for the enforceability of the *SAVM Act* is whether companies can adhere to the new laws. The Attorney-General, in the second reading speech, said that "[i]nternet platforms have the means to prevent the spread of abhorrent violent material".[104] Conversely, technology experts and industry heads are concerned that the Act has created a burden that is impossible to uphold with current technology. To illustrate the sheer amount of content which needs to be reviewed, over five hundred hours of video are uploaded to YouTube every minute,[105] and 350 million photos are uploaded every day on Facebook.[106]

Currently, most major social media companies used a hybrid regulation approach to review all content, employing a "people +

---

103 See *Corporations Act* 2001 (Cth) s 5(9).

104 Commonwealth, *Parliamentary Debates*, House of Representatives, 4 April 2019, 1850 (Christian Porter, Attorney General).

105 Google, 'Removals Under the Network Enforcement Law', *Transparency Report Google* (Web Page) <https://transparencyreport.google.com/netzdg/youtube?netzdg_examples=period:;type:;p:2&lu=netzdg_examples>.

106 Omnicore, 'Facebook by the Numbers: Stats, Demographics and Fun Facts', *Omnicore Agency* (Web Page, 10 February 2020) <https://www.omnicoreagency.com/facebook-statistics/>.

machine" framework.[107] For example, YouTube has four methods for reporting content which breach its guidelines:

1. *Automated matching by machine*: the use of technology to prevent re-uploads of known violative content, including the use of hashes, also known as digital fingerprints. Images and videos have unique hashes to prevent re-uploads of exact matches to videos removed for Community Guideline violations. There is also a shared industry database of hashes for certain content, such as child sexual abuse images, and terrorist recruitment videos.[108]
2. *Automated flagging by machine*: YouTube first used this machine learning technology in June 2017 to flag violent extremist content for human review.[109] The technology is trained to flag new content that might violate Community Guidelines using the database of videos which were already reviewed and removed for violent extremism by human reviewers. This technology should theoretically adapt and get smarter over time. The systems are most effective when there is a clearly defined target that is violative in any context. YouTube noted that this machine automation could not replace human judgement and nuance; it is used simply to flag content for further review.[110]
3. *Human flagging*: This reporting method is user-driven. The flagging system enables logged-in users to report content which potentially violets the Community Guidelines.[111]
4. *Legal complaints*: Following the German enactment of *NetzDG*, YouTube created additional reporting tools for individuals to report content that allegedly violates *NetzDG*. It has not yet been confirmed if similar measures will be taken for the *SAVM Act*.

Despite the multilayered approach taken for the identification and removal of content, the current framework is likely to be insufficient to remove all offending content 'expeditiously'. The proliferation of the footage from the Christchurch Massacre is a clear example of how the present technology is not always capable of preventing the proliferation of content which may be deemed AVM. First, Facebook AI did not flag the Christchurch Livestream. AI detection frameworks require 'training' using existing data to develop a model which is then used to flag future content. The AI tools did not detect the live stream of the

---

107 Google (n 105).
108 Ibid.
109 Ibid.
110 Ibid.
111 Ibid.

Massacre simply because there is not an extensive database of similar footage for the tools to be trained on.[112] There are two reasons for this, the first being that first-person footage of terrorist attacks is rare, second, if the algorithm flagged all first-person shooting content, it would also flag live-streaming gaming content.

For example, on Twitch, a platform dedicated to the live sharing of gaming, a Livestream of a shooting in 2018 remained accessible on the site for hours.[113]

Beyond the initial Livestream, Facebook, and all other social media services which had content depicting the Massacre would have been liable under the *SAVM Act* if they did not remove the content 'expeditiously'. After the initial video was flagged, Facebook immediately gave the video a hash so that the automated matching technology would catch copies at upload.[114] In the first 24 hours, more than 1.2 million videos of the attack were removed at upload, and 300,000 additional copies were removed after they were posted.[115] Despite the use of hashes and human review, seven weeks after the attack, there were still copies of the footage accessible on both Facebook and Instagram.[116] This is because users had edited, changed and manipulated the videos so that the original hash would not apply, this can include 'mirroring' the footage or embedding the footage in another video.[117] In only a few days, Facebook had found over 800 visually distinct variations of the video.[118]

It is clear that the social media and content providers, once alerted to the Livestream footage, were working to remove the videos expediently but the technology was simply unable to keep up. Facebook and major content providing and hosting services have taken steps to strengthen their algorithms; for example, in August 2019, Facebook open-sourced their algorithms for video and image detection.[119] There have been no significant developments or improvements in hashing or AI detection since the Christchurch Massacre.[120] The current AI

---

112 Evelyn Douek, 'Australia's "Abhorrent Violent Material' Law: Shouting "Nerd Harder" and Drowning Out Speech' 2020 94(41) *Australian Law Journal* 14.

113 Matthew Ingram, 'Twitch Joins the Unfortunate Social Media Club Where Death Happens in Real Time', *Colombia Journalism Review* (Web Page, 27 August 2018) <https://www.cjr.org/the_new_gatekeepers/twitch-shooting-livestream.php>.

114 Chris Sonderby, 'Update on New Zealand', *Facebook Newsroom* (Web Page, 18 March 2019) <https://about.fb.com/news/2019/03/update-on-new-zealand/>.

115 Rosen (n 3).

116 O'Sullivan (n 10).

117 Janis Dalins, Campbell Wilson and Douglas Boudry, 'PDQ & TMK + PDQF – A Test Drive of Facebook's Perceptual Hashing Algorithms', *Cornell University* <https://arxiv.org/abs/1912.07745> 7.

118 Sonderby (n 114).

119 Dalins, Wilson and Boudry (n 117) 1.

120 However, in February 2020, Facebook did release a white paper titled "Online Content Regulation", which can be found at <https://about.fb.com/wp-content/uploads/2020/02/Charting-A-Way-Forward_Online-Content-Regulation-White-Paper-1.pdf>.

technology and human-based frameworks are unable to remove all offending content after notification, even over several weeks.

Therefore, unless 'expeditiously' is defined to be a period of a few months, the *SAVM Act* has created a burden that is impossible for content and hosting service providers to uphold, drastically undermining the realistic enforceability of the offences. Further, attempts to adhere to the *SAVM Act* by social media providers may inadvertently result in adverse flow-on effects on the freedoms of expression and information.

The risk of the *SAVM Act* infringing on the international rights of freedom of expression and information arises not because the offences themselves unreasonably limit freedom of expression but rather because content and hosting services may over-censor in their efforts to avoid potential criminal liability. The 'expeditious' requirement poses a threat to the practical realisation of protection for freedom of expression and information in real-time.[121] As noted by the UN Special Rapporteur, the accelerated timelines imposed by the *SAVM Act* will not allow content and hosting services sufficient time to examine requests in detail, which may in practice mean that providers will consistently "produce an abundance of caution, for concern of financial fines and other consequences".[122] For example, during the aftermath of the Christchurch Massacre, for the first time, YouTube abandoned the human reviewing process. It instead removed all videos flagged as 'suspect', without waiting for human review. As aforementioned, AI technology is far from perfect. In addition to missing offending content, the AI may also incorrectly flag material which would not be classified as AVM. Even if the same machine + human framework is still employed, content reviewers will be under heightened pressure to make quick judgements about potentially offending content.

The uncertainty surrounding terms in the Act, such as the definition of 'reasonable time', or the complexity of underlying provisions of the *Criminal Code*, such as the definition of 'terrorist act', further heighten the insecurity providers feel under the new legislation. Further, although the defences protect content which is posted for purposes such as reporting or art, the time and effort required to make such nuanced assessments and preserve the protected exercises of freedom of expression conflict with the burden on service providers to 'expeditiously' remove content.[123] Presented with these conflicting considerations, it is probable the threat of criminal sanctions will tip the scales in favour of disproportionate restrictions on the freedom of expression and information, potentially undermining rather than

[121] Letter from Special Rapporteur (n 8) 5.
[122] Ibid.
[123] Ibid.

protecting the public interest.[124] This concern was also noted in Facebook's recent White Paper: "as to justify an approach that incentivises fast decision-making even where it may lead to more incorrect content removals (thus infringing on legitimate expression) or more aggressive detection regimes (such as in the case of using artificial intelligence to detect threats of self-harm, thus potentially affecting privacy interests)."[125]

A pertinent example is the potential impact on whistle-blowers. As noted by the Australian Law Council, "whistle-blowers may no longer be able to deploy social media to shine a light on atrocities committed around the world because social media companies will be required to remove certain content for fear of being charged with a crime".[126] For instance, during and following Syria's civil war, footage of the conflict was one of the only ways to prove that human rights violations occur.[127] Between 2012 and 2018 Google, using its machine learning algorithm, removed over 100,000 videos which depicted chemical weapon attacks that were carried out in Syria's civil war, destroying vital evidence of what took place.[128] Many of these videos were uploaded by a legitimate Douma-based news agency.[129]

Overall, under the current technology and frameworks, it is more or less impossible for content and hosting services to meet the removal and reporting requirements of the *SAVM Act* due to the shorter timeline. Further, the time pressure imposed by the offences inadvertently incentivises social media and hosting services to unreasonably restrict the freedoms of expression and information to avoid criminal sanctions.

VII Conclusion

Overall, the *SAVM Act*, while noble in its intentions, will face a variety of challenges which will hinder its ability to achieve its lofty goal of holding social media companies responsible for the content on their platforms. The *SAVM Act* contains a sufficient nexus to Australia to not be extraterritorial in its prescriptive jurisdiction, through the clauses

---

124 Ibid.

125 Monika Bickert, 'Charting a Way Forward: Online Content Regulation', *Facebook* (Article, July 2019) <https://about.fb.com/wp-content/uploads/2020/02/Charting-A-Way-Forward_Online-Content-Regulation-White-Paper-1.pdf>.

126 Commonwealth, *Parliamentary Debates*, House of Representatives, 4 April 2019, 1850 (Kerryn Phelps, MP).

127 Jon Porter, 'Upload Filters and One-Hour Takedowns: The EU's Latest Fight Against Terrorism Online, Explained', *The Verge* (Web Page, 21 March 2019) <https://www.theverge.com/2019/3/21/18274201/european-terrorist-content-regulation-extremist-terreg-upload-filter-one-hour-takedown-eu>.

128 Kate O'Flaherty, 'YouTube Keeps Deleting Evidence of Syrian Chemical Weapon Attacks', *Wired* (Web Page, 26 June 2018) <https://www.wired.co.uk/article/chemical-weapons-in-syria-youtube-algorithm-delete-video>.

129 Ibid.

stating that the content has to be suspected to be filmed in Australia, or is reasonably accessible in Australia, respectively. However, in its enforcement jurisdiction, the *SAVM Act* will face numerous issues as a result of its broad extraterritorial reach. First, there is a risk of conflicting international laws. This issue may arise in two ways: the requirements created by the *SAVM Act*, such as mandatory reporting, may require the content or hosting service to contravene a second state's law, or the enforcement of the *SAVM Act* may contravene a second state's laws, as illustrated in the *Yahoo! Case*. Second, the grey areas in international double jeopardy may cause conflict between the enforcement of the *SAVM Act*, and overlapping laws of other states, such as the *NetzDG* laws of Germany.

If the *SAVM Act* overcomes the challenges surrounding the extraterritorial reach, the Attorney-General may have difficulty choosing who to prosecute for breaches of the *SAVM Act*. The content the *SAVM Act* targets, unfortunately, often becomes viral content. This viral nature means that if the offending content is available on one platform, it will be accessible on numerous others as well. Prosecution of all platforms breaching the *SAVM Act* as a result of a singular incident would overwhelm the Australian court system. Therefore, the Attorney-General should target specific platforms.

Further, although at first glance, it would be simpler to charge locally registered subsidiaries of the larger companies, the subsidiaries tend to have minimal assets and not have the authority to change the moderation policies. Therefore, prosecuting the Australian subsidiaries would fail to achieve the underlying aim of the *SAVM Act*, to reduce the accessibility of AVM. Last, the majority of large social media providers outsource their content moderation to third parties, before prosecuting, the Attorney-General should consider the role, responsibility and potential liability of these content moderation services.

Finally, the *SAVM Act* creates a burden which is impossible to uphold with current AI technology. Most social media companies use a combination of human and machine-driven moderation techniques, including 'hashing' or 'digital fingerprinting'. However, even with these technologies, the platforms were unable to keep up with individuals sharing the Christchurch Massacre footage, which was still available weeks later. As noted by the Australian Law Council and the United Nations Special Rapporteur, the impossibly high bar set by the *SAVM Act* will likely result in over-policing by social media companies, resulting in potential breaches of the freedoms of speech and expression. The *SAVM Act* may have further negative, unintended, impacts on human rights. The obligations created may stifle whistle-blowers of international atrocities, or result in the destruction of crucial

evidence of human rights atrocities, hindering international prosecutions for crimes such as genocide.

Through the analysis of the potential challenges facing the *SAVM Act*, this article advises the Australian Government to consult legal societies, the United Nations, tech and social media companies, and human rights advocacy groups to address the many gaps in the legislation. For the *SAVM Act* to be enforceable and effective, the many questions raised by this article need to be considered. If these challenges are not adequately addressed, the *SAVM Act* will not only fail to achieve its goal of reducing the accessibility of abhorrent violent material; it will also pose a serious threat to the protection of human rights.